ON THE USE OF A MINIMAX REGRET FUNCTION TO SET SIGNIFICANCE POINTS IN PRIOR TESTS OF ESTIMATION

by

Richard J. Brook

Institute of Statistics
Mimeograph Series No. 856
Raleigh - December

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

1. QUADRATIC RISK FUNCTION FOR THE PREDICTED VALUE OF Y

1.1 Introduction and Review of Literature

When an applied statistician is faced with the problem of estimating certain parameters of interest, he has at least two choices before him. From the data at his disposal, he can use the method of ordinary least squares to obtain one set of estimates. Another set of estimates can be found by the method of restricted least squares. In the latter case, the constraints may take the form of pooling two sets of data. A different kind of restriction is the use of prior knowledge to estimate certain components of the $\underset{\sim}{\beta}$ vector, or linear combinations of these. All these constraints can be conveniently written as $H\underset{\sim}{\beta} = \underset{\sim}{h}$, where H is an mxk matrix of known constants, $\underset{\sim}{\beta}$ is a kxl vector of unknown parameters, and $\underset{\sim}{h}$ is a kxl vector of known constants.

This formulation includes the following common restrictions:

(i) One parameter is a constant multiple of another; e.g.,

$$\beta_1 = 3\beta_2 .$$

(ii) One parameter has a certain constant value; e.g.,

$$\beta_1 = 2 .$$

(iii) A linear combination of some of the parameters has a constant value; e.g.,

$$\beta_1 + 2\beta_2 = 0 .$$

(iv) Wallace[1] points out that the pooling problem is equivalent to choosing H to be $(I_k, - I_k)$, where I_k is the kxk identity matrix, and there are k components of the pooled $\underset{\sim}{\beta}$ vector.

[1]T. D. Wallace, Professor, N. C. State University, Raleigh, N. C., Weaker criteria and tests for linear restrictions, to be published in Econometrica.

Theil and Goldberger [1961] indicate how certain a priori inequality constraints on the $\underset{\sim}{\beta}$ vector can also be included in this formulation.

In practice, it is not known whether the constraints are exact, so that the experimenter must decide whether to use the ordinary least squares estimate, $\underset{\sim}{b}$, or the restricted estimator, $\hat{\underset{\sim}{\beta}}$. The data, itself, can be used to decide which set of estimates to use by means of a prior test of estimation. The test statistic takes the form,

$u = \dfrac{SSE(\hat{\underset{\sim}{\beta}}) - SSE(\underset{\sim}{b})}{k\ SSE(\underset{\sim}{b})}$ where SSE( ) signifies the sum of squares for error using that estimate of $\underset{\sim}{\beta}$, and k is a constant to make u the ratio of two mean squares.

A sequential estimator of $\underset{\sim}{\beta}$ can now be defined as

$$\underset{\sim}{\beta}^* = \begin{cases} \underset{\sim}{b} \text{ if } u \geq \lambda \\ \hat{\underset{\sim}{\beta}} \text{ if } u < \lambda \end{cases}$$

where λ is an appropriate constant, and, generally, will be a certain critical value of Snedecor's F distribution.

It is well known that $\underset{\sim}{\beta}^*$ is a biased estimator. In Chapter 4, its bias and quadratic risk function will be evaluated. Similarly, $\underset{\sim}{y}^* = X\underset{\sim}{\beta}^*$ is a sequential estimator for the predicted value of $\underset{\sim}{y}$. In the remaining sections of Chapter 1, the bias and quadratic risk function for $\underset{\sim}{y}^*$ will be evaluated. To do this, a method will be used similar to that employed by Larson and Bancroft [1963], although they considered the simpler case of zero restrictions instead of general linear restrictions. Ashar [1970] also evaluated the bias and mean square error for a sequential estimator for a two variable model with zero restrictions.

In the definition of $\underset{\sim}{\beta}^*$, or $X\underset{\sim}{\beta}^*$, there is a certain arbitrariness as λ is chosen at will by the experimenter. Sawa and Hiromatsu [1971], in a similar context, defined a minimax regret function in order to specify a value for λ. Their general approach will be used to define a regret function for $\underset{\sim}{y}^*$ in Chapter 3, and for $\underset{\sim}{\beta}^*$ in Chapter 4.

Toro-Vizcarrondo and Wallace [1968] and Wallace[2] proposed other criteria than testing whether the linear restrictions are true, which is equivalent to testing that the non-centrality parameter, θ, is zero. Their criteria lead to testing whether θ is less than 1/2, or m/2, where m is the number of restrictions. During this study, comparisons will be made with the results they obtained.

1.2 The Linear Model and Restrictions Defined

Consider the general linear model $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$, where $\underset{\sim}{y}$, $\underset{\sim}{\beta}$ and $\underset{\sim}{\varepsilon}$ are nx1 vectors and X is an nxk matrix of fixed constants. n is assumed larger than k, and X has rank k, that is, of full rank. It is not necessary that the column vectors of X be orthogonal. The error vector $\underset{\sim}{\varepsilon}$ is assumed to be distributed as a multivariate normal with mean zero and variance-covariance matrix $\sigma^2 I$. That is, the errors ε_i, i = 1, 2, 3, ..., n, are identically and independently distributed as $N(0, \sigma^2)$.

The ordinary least squares estimate (o.l.s.) of $\underset{\sim}{\beta}$ is $\underset{\sim}{b} = S^{-1}X'\underset{\sim}{y}$, where $S = X'X$.

If a set of exact linear restrictions are placed on the $\underset{\sim}{\beta}$ vector, viz., $H\underset{\sim}{\beta}=\underset{\sim}{h}$ where H is an mxk matrix of rank m, and $\underset{\sim}{h}$ is an mx1 vector, then the restricted estimator of $\underset{\sim}{\beta}$ can be found by the method of Lagrange multipliers, as in Goldberger [1964, p. 256].

---

[2] See footnote 1.

The expression $Z = (\underset{\sim}{y} - X\underset{\sim}{\beta})'(\underset{\sim}{y} - X\underset{\sim}{\beta}) - 2\mu'(H\underset{\sim}{\beta}-\underset{\sim}{h})$ is minimized with respect to $\underset{\sim}{\beta}$ and $\underset{\sim}{\mu}$, the mx1 vector of Lagrange multipliers. The restricted estimator, $\hat{\underset{\sim}{\beta}}$, for $\underset{\sim}{\beta}$ is found to be

$$\hat{\underset{\sim}{\beta}} = \underset{\sim}{b} - S^{-1}H'[HS^{-1}H']^{-1}[H\underset{\sim}{b}-\underset{\sim}{h}] \; . \tag{1.2.1}$$

The distribution of these two estimators, $\underset{\sim}{b}$ and $\hat{\underset{\sim}{\beta}}$, will be needed in the next section. Clearly,

$$\underset{\sim}{b} \sim N(\underset{\sim}{\beta},\; S^{-1}\sigma^2) \quad . \tag{1.2.2}$$

As H, S, and $\underset{\sim}{h}$ consist of fixed constants, then

$$\begin{aligned} E\hat{\underset{\sim}{\beta}} &= E\{\underset{\sim}{b} - S^{-1}H'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-\underset{\sim}{h})\} \\ &= \underset{\sim}{\beta} - S^{-1}H'[HS^{-1}H']^{-1}(H\underset{\sim}{\beta}-\underset{\sim}{h}) \; . \end{aligned} \tag{1.2.3}$$

The variance of $\hat{\underset{\sim}{\beta}}$ is then

$$= E\{[(\underset{\sim}{b}-\underset{\sim}{\beta}) - S^{-1}H'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-H\underset{\sim}{\beta})][(\underset{\sim}{b}-\underset{\sim}{\beta}) - S^{-1}H'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-H\underset{\sim}{\beta})]'\}.$$

On simplifying, this becomes

$$\text{var}\; \hat{\underset{\sim}{\beta}} = \sigma^2[S^{-1} - S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}] \quad \text{and}$$

$$\hat{\underset{\sim}{\beta}} \sim N\;(\underset{\sim}{\beta} - S^{-1}H'[HS^{-1}H']^{-1}(H\underset{\sim}{\beta}-\underset{\sim}{h}) \;,\; \text{var}\; \hat{\underset{\sim}{\beta}}\;) \tag{1.2.4}$$

where the var $\hat{\underset{\sim}{\beta}}$ is given above.

If it is known that the linear restrictions are exact, then $\hat{\underset{\sim}{\beta}}$ should be chosen in preference to $\underset{\sim}{b}$, as both estimators would be unbiased, and it can be shown that

$$\text{var}\; \hat{\underset{\sim}{\beta}} \leq \text{var}\; \underset{\sim}{b} \quad . \tag{1.2.5}$$

This follows from the fact that the matrix $S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}$ is non-negative definite.

It should be noted that (1.2.5) holds whether or not the restrictions are exact.

1.3 The Preliminary Test of Estimation

In this section, a few comments will be given on the preliminary test of estimation and the distributions of some important statistics will be derived as these distributions will be needed in later sections.

If it is not known whether the linear restrictions are exact, then a prior test could be performed, namely testing the null hypothesis,

$$H_o: \; H\underset{\sim}{\beta} = \underset{\sim}{h} \quad .$$

Rao [1965] shows that a U.M.P. test of this hypothesis is based on the statistic

$$u = \frac{SSE(\hat{\underset{\sim}{\beta}}) - SSE(\underset{\sim}{b})}{m} \div \frac{SSE(\underset{\sim}{b})}{n-k} \quad . \tag{1.3.1}$$

This expression can be simplified by substituting the value of $\hat{\underset{\sim}{\beta}}$ found in (1.2.1). The resulting expression for u can be written as

$$u = \frac{Q}{m\hat{\sigma}^2} \quad \text{where } \hat{\sigma}^2 = \frac{SSE(\underset{\sim}{b})}{n-k} \quad , \tag{1.3.2}$$

that is, the sample estimate for σ^2 with n-k degrees of freedom, and

$$Q = (H\underset{\sim}{b}-\underset{\sim}{h})'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-\underset{\sim}{h}) \quad . \tag{1.3.3}$$

Now, from (1.2.2) $\underset{\sim}{b} \sim N(\underset{\sim}{\beta}, S^{-1}\sigma^2)$ and consequently

$$(H\underset{\sim}{b}-\underset{\sim}{h}) \sim N\,(H\underset{\sim}{\beta}-\underset{\sim}{h}, HS^{-1}H'\sigma^2) \quad . \tag{1.3.4}$$

As H was assumed to be an mxk matrix of rank m, and S is a kxk matrix of rank k, it follows that $HS^{-1}H'$ is an mxm matrix of rank m. Graybill [1961, p. 84], shows that this is a sufficient condition, along with (1.3.4), to claim that:

$$\frac{Q}{\sigma^2} = \frac{(H\underset{\sim}{b}-\underset{\sim}{h})'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-\underset{\sim}{h})}{\sigma^2} \tag{1.3.5}$$

is distributed as a non-central χ^2 with parameters m and θ which is usually written as $\chi'^2(m;\theta)$, where θ is the non-centrality parameter and

$$\theta = \frac{(H\underset{\sim}{\beta}-\underset{\sim}{h})'[HS^{-1}H']^{-1}(H\underset{\sim}{\beta}-\underset{\sim}{h})}{2\sigma^2} . \tag{1.3.6}$$

Under the null hypothesis, the distribution of $\frac{Q}{\sigma^2}$ reduces to a central χ^2 with m degrees of freedom, that is $\chi^2(m)$. It is well known that

$$V = \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k) . \tag{1.3.7}$$

In section 1.5, use will be made of the fact that $\underset{\sim}{b}$ and V are independently distributed, and also that V and Q are independent sums of squares. It will be useful to justify these two statements at this point.

The first fact is well known and is proven in detail in Graybill [1961, p. 113]. In brief, $\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y}$ is a linear form in $\underset{\sim}{y}$, and

$$V = \frac{1}{\sigma^2}[\underset{\sim}{y}-X\underset{\sim}{b}]'[\underset{\sim}{y}-X\underset{\sim}{b}], \quad \text{or}$$

$$V = \frac{1}{\sigma^2}\underset{\sim}{y}'[I - XS^{-1}X']\underset{\sim}{y} \tag{1.3.8}$$

which is a quadratic form in $\underset{\sim}{y}$. Independence of $\underset{\sim}{b}$ and V follows from the fact that $\{(X'X)^{-1}X'\}\{I - XS^{-1}X'\} = 0$ which satisfies the conditions of a theorem quoted by Graybill [1961, p. 87].

To show the independence of V and Q, (1.3.8) can be rewritten, substituting $X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}$ for $\underset{\sim}{y}$, to give

$$V = \frac{1}{\sigma^2}\underset{\sim}{\varepsilon}'[I - XS^{-1}X']\underset{\sim}{\varepsilon} \quad . \tag{1.3.9}$$

From (1.3.3), $\quad Q = (H\underset{\sim}{b}-\underset{\sim}{h})'[HS^{-1}H']^{-1}(H\underset{\sim}{b}-\underset{\sim}{h}) \quad .$

Now, $\underset{\sim}{b} = (X'X)^{-1}X'\underset{\sim}{y} = (X'X)^{-1}X'(X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}) = \underset{\sim}{\beta} + S^{-1}X'\underset{\sim}{\varepsilon}$. Substituting this value for $\underset{\sim}{b}$ in (1.3.3), and recalling that under the null hypothesis, $H\underset{\sim}{\beta}=\underset{\sim}{h}$, then, under the null hypothesis, Q reduces to the following central quadratic form:

$$Q = \underset{\sim}{\varepsilon}'\{XS^{-1}H'[HS^{-1}H']^{-1}HS^{-1}X'\}\underset{\sim}{\varepsilon} \quad . \tag{1.3.10}$$

Johnson and Kotz [1970, p. 176] point out that to show the independence of quadratic forms only the central forms need to be considered. The necessary and sufficient condition for V and Q to be independent is that the product of the matrices $[I - XS^{-1}X']$ from (1.3.9), and $\{XS^{-1}H'[HS^{-1}H']^{-1}HS^{-1}X'\}$ from (1.3.10) be zero. By straightforward multiplication, this is seen to be true. Two additional facts will prove useful in section 1.5. As $\underset{\sim}{b}$ and V are independent, and $\underset{\sim}{b} \sim N(\underset{\sim}{\beta}, S^{-1}\sigma^2)$, then the joint density of $\underset{\sim}{b}$ and V can be written as

$$\begin{aligned} f(\underset{\sim}{b}, V) &= f_1(\underset{\sim}{b})\, f_2(V) \qquad (1.3.11)\\ &= K\, f_2(V)\, \exp\left[\frac{-(\underset{\sim}{b}-\underset{\sim}{\beta})'S(\underset{\sim}{b}-\underset{\sim}{\beta})}{2\sigma^2}\right] . \end{aligned}$$

As Q and $V\sigma^2 = (n-k)\hat{\sigma}^2$ are independent sums of squares, then

$$u = \frac{Q}{m\hat{\sigma}^2} \tag{1.3.12}$$

is distributed as the non-central F distribution with parameters m, n-k and θ.

1.4 The Estimator of y^*

In this section, the common estimator for the predicted value of y,[3] y^*, based on preliminary tests of estimation, will be considered. Its bias and mean square error will be evaluated. y^* will be taken to be

$$X\beta^* = \begin{cases} Xb \text{ if } u \geq \lambda \\ X\hat{\beta} \text{ if } u < \lambda \end{cases} \tag{1.4.1}$$

where λ is the critical point of Snedecor's F statistic with m and n-k degrees of freedom at the desired type I error level. This could also be written as

$$y^* = X\beta^* = X\hat{\beta}\, 1_{[0,\lambda)}(u) + Xb\, 1_{[\lambda,\infty)}(u)$$

where $1_{[a,b)}(u)$, the characteristic function, $= \begin{cases} 1 \text{ if u is in } [a,b) \\ 0 \text{ if u is in } [a,b)^c \end{cases}$.

In terms of the null hypothesis, H_o: $H\beta=h$, the estimator $X\hat{\beta}$ is chosen if H_o is accepted, but Xb is chosen if H_o is rejected at the desired α-level.

[3]From now on the ($\sim$) will be omitted from $\underset{\sim}{y}$, $\underset{\sim}{b}$, $\underset{\sim}{\hat{\beta}}$, $\underset{\sim}{\beta^*}$, $\underset{\sim}{\beta}$ but symbols y, b, $\hat{\beta}$, β^*, β will still represent vectors.

1.5 The Bias Function of y^*

Let A_1 be the set $\{ (b,V); \frac{(Hb-h)'[HS^{-1}H']^{-1}(Hb-h)}{m\hat{\sigma}^2} \geq \lambda \}$ and A_o be the complement of A_1, that is,

$$A_o = \{ (b,V) ; \frac{(Hb-h)'[HS^{-1}H']^{-1}[Hb-h]}{m\hat{\sigma}^2} < \lambda \} .$$

The probability of the set A_1 is given by,

$$P(A_1) = \iint_{A_1} K f_2(V) \exp \left(\frac{-(b-\beta)'S(b-\beta)}{2\sigma^2} \right) dbdV \qquad (1.5.1)$$

and by transforming the variables b,V to u as defined by (1.3.2)

$$P(A_1) = \int_{u \geq \lambda} f_u(u;m,n-k,\theta)du = g(\theta,\lambda) \qquad (1.5.2)$$

where $f_u(u;m,n-k,\theta)$ is the density of the non-central F with m and n-k degrees of freedom, that is, $u \sim F'(m,n-k;\theta)$ as was shown in (1.3.12). However, rather than using the density function $f_u(u;m,n-k,\theta)$, a transformation $t = \frac{m\ u}{n-k} = \frac{Q}{(n-k)\hat{\sigma}^2}$ will be employed. The resulting density $f_t(t;m,n-k,\theta)$ is the ratio of two independent sums of squares.

Kempthorne [1967, p. 221] shows that

$$f_t(t;m,n-k,\theta) = \sum_{i=o}^{\infty} \frac{\theta^i e^{-\theta} t^{i+\frac{m}{2}-1}}{i!\ B(\frac{m}{2}+i, \frac{n-k}{2})\ \ (1+t)^{i+\frac{m+n-k}{2}}} \qquad (1.5.3)$$

where B(p,q) is the Beta function with parameters p and q. Now, the expected value of y^* can be considered separately over the sets A_o and A_1 for these form a partition of the parameter space.

$$E(y^*) = E(X\beta^*|A_o)P(A_o) + E(X\beta^*|A_1)P(A_1) \qquad (1.5.4)$$

$$= E(X\hat{\beta}|A_o)P(A_o) + E(Xb|A_1)P(A_1) \quad .$$

In (1.2.1), $\hat{\beta}$ was shown to equal $b - S^{-1}H'[HS^{-1}H']^{-1}(Hb-h)$. As b is an unbiased estimate of β, then

$$E(Xb|A_o)P(A_o) + E(Xb|A_1)P(A_1) = E(Xb) = X\beta \quad . \qquad (1.5.5)$$

$E(y^*)$ reduces to:

$$E(y^*) = X\beta - E(XS^{-1}H'[HS^{-1}H']^{-1}(Hb-h)|A_o)P(A_o) \quad . \qquad (1.5.6)$$

To evaluate this expression, it is necessary to find $E(b|A_o)P(A_o)$. For convenience, $E(b|A_1)P(A_1)$ will be evaluated first. One way of finding this is to differentiate (1.5.1) and (1.5.2) with respect to the β vector. Theil [1971, pp. 30-31] gives a review of vector and matrix differentiation. From his comments there, it is clear that

$$\frac{\partial}{\partial\beta}[(b-\beta)'S(b-\beta)] = -(S+S')(b-\beta) = -2S(b-\beta) \quad , \qquad (1.5.7)$$

as S is symmetric.

From (1.5.1) and (1.5.7)

$$\frac{\partial P(A_1)}{\partial\beta} = \iint_{A_1} K\, f_2(V)\, \frac{S(b-\beta)}{\sigma^2} \exp\left(\frac{-(b-\beta)'S(b-\beta)}{2\sigma^2}\right) dbdV$$

$$= E\left(\frac{S(b-\beta)}{\sigma^2}\,\middle|\,A_1\right)P(A_1) \quad . \qquad (1.5.8)$$

But from (1.5.2),

$$\frac{\partial P(A_1)}{\partial \beta} = \frac{\partial g(\theta,\lambda)}{\partial \beta} = g'(\theta,\lambda)\,\frac{\partial \theta}{\partial \beta} \quad . \tag{1.5.9}$$

From (1.3.6), $\theta = \dfrac{(H\beta-h)'[HS^{-1}H']^{-1}(H\beta-h)}{2\sigma^2}$, so that

$$\frac{\partial \theta}{\partial \beta} = \frac{H'[HS^{-1}H']^{-1}(H\beta-h)}{\sigma^2} \quad . \tag{1.5.10}$$

From (1.5.8) and (1.5.10),

$$E\left(\frac{S(b-\beta)}{\sigma^2}\,\Big|A_1\right)P(A_1) = \frac{H'[HS^{-1}H']^{-1}(H\beta-h)}{\sigma^2}\,g'(\theta,\lambda) \ .$$

This expression can be premultiplied by S^{-1}, as S is a full-rank matrix of constants. Thus,

$$E(b-\beta|A_1)P(A_1) = S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)g'(\theta,\lambda) \quad , \tag{1.5.11}$$

and, as β is a vector of constants, $E(\beta|A_1)P(A_1) = \beta\ g\ (\theta,\lambda)$. Hence,

$$E(b|A_1)P(A_1) = S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)g'(\theta,\lambda) + \beta g\ (\theta,\lambda) \ . \tag{1.5.12}$$

From the unbiasedness of b, it then follows that

$$E(b|A_o)P(A_o) = \beta(1-g(\theta,\lambda)) - S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)g'(\theta,\lambda) \ . \tag{1.5.13}$$

As the bias of $y^* = E(y^*) - X\beta$, then from (1.5.6) and (1.5.13), and bearing in mind that H, S and h are fixed, then

$$\begin{aligned}
\text{bias } y^* = & - [\, XS^{-1}H'[HS^{-1}H']^{-1}H\ E(b|A_o)P(A_o) \\
& - XS^{-1}H'[HS^{-1}H']^{-1}h\ P(A_o)\ ] \\
= & - XS^{-1}H'[HS^{-1}H']^{-1}H\ \beta(1-g(\theta,\lambda)) \\
& + XS^{-1}H'[HS^{-1}H']^{-1}HS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)g'(\theta,\lambda) \\
& + XS^{-1}H'[HS^{-1}H']^{-1}h(1-g(\theta,\lambda))
\end{aligned} \qquad (1.5.14)$$

On simplifying this leads to

$$\text{bias } y^* = XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)\{g'(\theta,\lambda) + g(\theta,\lambda)-1\}. \qquad (1.5.15)$$

This expression could be simplified by evaluating $g'(\theta,\lambda)$. From (1.5.2) and (1.5.3), $g(\theta,\lambda)$ can be written as

$$g(\theta,\lambda) = \int_{u\geq\lambda} f_u(u;m,n-k,\theta)du = \int_{t>\frac{m\lambda}{n-k}} f_t(t;m,n-k,\theta)dt. \qquad (1.5.16)$$

This equation can be differentiated with respect to θ. To do this, it would be convenient if the last expression could be differentiated under the integral sign. Bartle [1964, p. 307] proves that necessary conditions for this are that $f_t(t;m,n-k,\theta)$ and $\frac{\partial}{\partial\theta} f_t(t;m,n-k,\theta)$ be continuous in θ. From (1.5.3) and (1.5.17) and (1.5.18) below, it is clear that these conditions are met. Now

$$\frac{\partial}{\partial\theta} f_t(t;m,n-k,\theta) = -f_t(t;m,n-k,\theta) + \sum_{i=1}^{\infty} \frac{i\ \theta^{i-1}\ e^{-\theta}\ t^{i+\frac{m}{2}-1}}{i!\ B(\frac{m}{2}+i,\frac{n-k}{2})\ (1+t)^{i+\frac{m+n-k}{2}}}.$$

The second term on the right of (1.5.17) can be simplified by putting $j=i-1$.

$$\sum_{i=1}^{\infty} \frac{\theta^{i-1}\, e^{-\theta}\; t^{\,i+\frac{m}{2}-1}}{(i-1)!\; B(\frac{m}{2}+i,\ \frac{n-k}{2})\quad (1+t)^{i+\frac{m+n-k}{2}}} \tag{1.5.18}$$

$$= \sum_{j=o}^{\infty} \frac{\theta^{j}\, e^{-\theta}\; t^{\,j+1+\frac{m}{2}-1}}{j!\; B(\frac{m}{2}+1+j,\frac{n-k}{2})\cdot(1+t)^{j+1+\frac{m+n-k}{2}}}$$

$$= \sum_{j=o}^{\infty} \frac{\theta^{j}\, e^{-\theta}\; t^{\,j+\frac{m^*}{2}-1}}{j!\; B(\frac{m^*}{2}+j,\ \frac{n-k}{2})\ (1+t)^{j+\frac{m^*+n-k}{2}}}\ , \text{ where } m^* = m+2$$

$= \quad f_t(t;m+2,n-k,\theta)$, and this is the density of the ratio of two sums of squares with m+2 degrees of freedom in the numerator and (n-k) degrees of freedom in the denominator.

Let $r(\theta,\lambda) = \int_{t \geq \frac{m\lambda}{n-k}} f_t(t;m+2,n-k,\theta)\ dt$

$$= P\ \{\frac{m+2}{n-k}\ F'(m+2,n-k;\theta) \geq \frac{m\lambda}{n-k}\}\ .$$

In comparison,

$$g(\theta,\lambda) = P\ \{\frac{m}{n-k}\ F'\ (m,n-k;\theta) \geq \frac{m\lambda}{n-k}\}$$

with this definition of $r(\theta,\lambda)$,

$$g'(\theta,\lambda) = -g(\theta,\lambda) + r(\theta,\lambda) \tag{1.5.19}$$

and

$$\text{Bias } y^* = XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(r(\theta,\lambda)-1)\ . \tag{1.5.20}$$

As could be expected, the bias function is zero if the conditions $H\beta=h$ are exact, and the absolute value of the bias function increases as $|H\beta-h|$ increases, that is, to the extent that β does not satisfy the restrictions. From (1.2.4), $\hat{\beta} \sim N(\beta-S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h), \Sigma_{\hat{\beta}\hat{\beta}})$, so that $|\text{bias } X\hat{\beta}| = |-XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)|$. This leads to

$$|\text{bias } y^*| = |\text{bias } X\hat{\beta}| \quad |1-r(\theta,\lambda)| \leq |\text{bias } X\hat{\beta}| \tag{1.5.21}$$

as $r(\theta,\lambda)$, being a probability, lies between 0 and 1.

As Xb is an unbiased estimator of y^*, then the value of $|\text{bias } y^*|$ lies between that of $|\text{bias } Xb|$ and $|\text{bias } X\hat{\beta}|$. That is,

$$|\text{bias } Xb| \leq |\text{bias } y^*| \leq |\text{bias } X\hat{\beta}| \quad . \tag{1.5.22}$$

Now the estimator Xb will be chosen with probability $P(A_1) = g(\theta,\lambda)$ and $X\hat{\beta}$ with probability $P(A_0) = 1-g(\theta,\lambda)$. It may be thought, naively, that the $|\text{bias } y^*|$ would be given by the expression

$$\begin{aligned} |\text{bias } y^*| &= |\text{bias } Xb|g(\theta,\lambda) + |\text{bias } X\hat{\beta}|(1-g(\theta,\lambda)) \\ &= |\text{bias } X\hat{\beta}| \ (1-g(\theta,\lambda)) \quad . \end{aligned}$$

For θ and λ in the open interval $(0, +\infty)$, this expression is actually larger than the correct expression of (1.5.21). This can be verified by employing an analogous argument to that which will be used later in Lemma 3.2.2 to show that

$$r(\theta,\lambda) > g(\theta,\lambda) \quad \text{for } \theta,\lambda \quad \text{in } (0, +\infty),$$

and hence that

$$1 - g(\theta,\lambda) \quad > \quad 1 - r(\theta,\lambda) \quad .$$

It is not clear from the expression for bias y^* in (1.5.20) whether the absolute value of this bias function increases or decreases as θ, the non-centrality parameter increases without limit. For large θ, $|H\beta-h|$ increases but $r(\theta,\lambda)$ tends to 1 as Lemma 3.2.3 will demonstrate. This problem can be posed in another way by considering (bias y^*)$'$(bias y^*).

$$(\text{bias } y^*)'(\text{bias } y^*) = (H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}X'XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)[r(\theta,\lambda)-1]^2$$

$$= 2\sigma^2\,\theta[r(\theta,\lambda)-1]^2 \quad . \tag{1.5.23}$$

The problem reduces as to whether θ goes to infinity faster than $[r(\theta,\lambda)-1]^2$ goes to zero.

As Lemma 3.2.7 will show more formally, $\theta[r(\theta,\lambda)-1] \to 0$ as $\theta \to +\infty$. Also, $r(\theta,\lambda)$ is an increasing function of θ so that $r(\theta,\lambda)-1 \to 0$ as $\theta \to +\infty$. Consequently, $\theta[r(\theta,\lambda)-1]^2$ and hence the quadratic form, (bias y^*)$'$(bias y^*) $\to 0$ as $\theta \to +\infty$, which, in turn, implies that bias $y^* \to 0$ as $\theta \to +\infty$.

1.6 The Quadratic Risk Function of y^*

The mean square error of $y^* = E(y^* - X\beta)(y^* - X\beta)'$. Instead of the mean square error, however, its trace will be evaluated and, in the third chapter, this will be used to define a minimax regret function.

The problem, then, is to evaluate $M(\theta,\lambda) = E(X\beta^*-X\beta)'(X\beta^*-X\beta)$. The motivation for using $M(\theta,\lambda)$ is that this is the quadratic risk function of y^*. Thus,

$$\begin{aligned} M(\theta,\lambda) &= E(X\beta^*-X\beta)'(X\beta^*-X\beta) \qquad (1.6.1)\\ &= E[(Xb-X\beta)'(Xb-X\beta)|A_1]P(A_1)\\ &+ E[(X\hat\beta-X\beta)'(X\hat\beta-X\beta)|A_o]P(A_o)\\ &= E[(b-\beta)'S(b-\beta)|A_1]P(A_1)\\ &+ E[(\hat\beta-\beta)'S(\hat\beta-\beta)|A_o]P(A_o) \end{aligned}$$

Noting the form of $\hat{\beta}$ from (1.2.1), then

$$M(\theta,\lambda) = E[(b-\beta)'S(b-\beta)|A_1]P(A_1) + E[(b-\beta-S^{-1}H'[HS^{-1}H']^{-1}[Hb-h])'S(b-\beta-S^{-1}H'[HS^{-1}H']^{-1}[Hb-h]|A_o)P(A_o) \quad (1.6.2)$$

Expanding the last term on the right of (1.6.2), there are four terms under the expected value:

(a) $(b-\beta)'S(b-\beta)$

(b) $-[Hb-h]'[HS^{-1}H']^{-1}HS^{-1}S(b-\beta)$

(c) $-(b-\beta)'S\ S^{-1}H'[HS^{-1}H']^{-1}[Hb-h]$

(d) $[Hb-h]'[HS^{-1}H']^{-1}HS^{-1}S\ S^{-1}H'[HS^{-1}H']^{-1}[Hb-h]$

Now (d) reduces to $[Hb-h]'[HS^{-1}H']^{-1}[Hb-h]$ which is seen to be the Q defined in Section 1.3.

As $H(b-\beta) = (Hb-h) - (H\beta-h)$, (b) can be written as

$$- [Hb-h]'[HS^{-1}H']^{-1}[(Hb-h) - (H\beta-h)]$$

$$= - Q + [Hb-h]'[HS^{-1}H']^{-1}[H\beta-h]$$

(c) is merely the transpose of (b).

Now,

$$E[(b-\beta)'S(b-\beta)|A_o]P(A_o) + E[(b-\beta)'S(b-\beta)|A_1]P(A_1) \quad (1.6.3)$$

$$= E[(b-\beta)'S(b-\beta)]$$

$= k\ \sigma^2$ from the comment (1.2.2) and the fact that S is a kxk matrix of rank k.

Also, $Q = (Hb-h)'[HS^{-1}H']^{-1}(Hb-h)$ is symmetric, so that $Q = Q'$.
(1.6.2) can be written

$$M(\theta,\lambda) = E[(b-\beta)'S(b-\beta)|A_1]P(A_1) + E[(b-\beta)'S(b-\beta)|A_o]P(A_o) \quad (1.6.4)$$

$$+ E[-Q + (Hb-h)'[HS^{-1}H']^{-1}(H\beta-h)|A_o]P(A_o)$$

$$+ E[-Q' + (H\beta-h)'[HS^{-1}H']^{-1}(Hb-h)|A_o]P(A_o)$$

$$+ E[Q|A_o]P(A_o) .$$

With the above simplifications (1.6.4) becomes:

$$M(\theta,\lambda) = k\ \sigma^2 + E[(Hb-h)'[HS^{-1}H']^{-1}(H\beta-h)|A_o]P(A_o) \quad (1.6.5)$$

$$+ E[(H\beta-h)'[HS^{-1}H']^{-1}(Hb-h)|A_o]P(A_o) - E(Q|A_o)P(A_o) .$$

In (1.5.13) an expression was obtained for $E(b|A_o)P(A_o)$, so that it only remains to evaluate $E(Q|A_o)P(A_o)$. From (1.5.16)

$$E(Q|A_1)P(A_1) = \int_{u \geq \lambda} Q\ f_u(u;\theta)du .$$

As $u = \frac{Q}{m\hat{\sigma}^2}$ and it was shown in (1.3.6) that Q and $\hat{\sigma}^2$ are independent, and for convenience, let $q = \frac{Q}{\sigma^2}$, then

$$E(q|A_1)P(A_1) = \int_o^\infty \int_{q \geq \frac{m\lambda\hat{\sigma}^2}{\sigma^2}} q\ f_1(q;m,\theta)f_2(\hat{\sigma}^2)dqd\hat{\sigma}^2 \quad (1.6.6)$$

where $f_1(.,.)$, $f_2(.)$ are the respective marginal densities of q and $\hat{\sigma}^2$. From (1.3.2), $f_1(q;m,\theta)$ is a non-central χ'^2 density function with m degrees of freedom and with non-centrality parameter θ.

Consider then $f_1(q;m,\theta) = e^{-\theta} \sum_{i=0}^{\infty} \frac{\theta^i}{i!} \frac{e^{-\frac{q}{2}} \; q^{i+\frac{m}{2}-1}}{2^{i+\frac{m}{2}} \; \Gamma(i+\frac{m}{2})}$ where $\Gamma(\;)$ is the gamma function.

Multiplying this density by q can be thought of as increasing the value of m by 2 in the exponent of q. In the denominator, m can also be increased by 2 if the whole expression under the summation sign is multiplied by a factor $2[i+\frac{m}{2}]$. This gives

$$qf_1(q;m,\theta) = e^{-\theta} \sum_{i=0}^{\infty} \frac{\theta^i \; e^{-\frac{q}{2}} \; q^{i+\frac{m+2}{2}-1}}{i! \; 2^{i+\frac{m+2}{2}} \quad \Gamma(i+\frac{m+2}{2})} \; 2[i+\frac{m}{2}] \qquad (1.6.7)$$

$$= 2\theta \; e^{-\theta} \sum_{i=0}^{\infty} \frac{i\theta^{i-1} \; e^{-\frac{q}{2}} \; q^{i+\frac{m+2}{2}-1}}{i! \; 2^{i+\frac{m+2}{2}} \quad \Gamma(i+\frac{m+2}{2})}$$

$$+ m \; e^{-\theta} \sum_{i=0}^{\infty} \frac{\theta^i \; e^{-\frac{q}{2}} \; q^{i+\frac{m+2}{2}-1}}{i! \; 2^{i+\frac{m+2}{2}} \quad \Gamma(i+\frac{m+2}{2})} .$$

The last term on the right-hand side of the expression (1.6.7) is obviously m $f_1(q;m+2,\theta)$, and if in the first term j is put equal to i-1, then in a method similar to that explained in (1.5.18), the first term becomes $2\theta \; f_1(q;m+4,\theta)$. Putting these results back in (1.6.6) gives

$$E(q|A_1)P(A_1) = \int_0^{\infty} \int_{q \geq \frac{m\lambda\hat{\sigma}^2}{\sigma^2}} \{2\theta \; f_1(q;m+4,\theta) + m \; f_1(q;m+2,\theta)\} \; f_2(\hat{\sigma}^2) dq d\hat{\sigma}^2 . \qquad (1.6.8)$$

In section 1.5, the statistic $t = \frac{Q}{(n-k)\hat{\sigma}^2} = \frac{q}{(n-k)\hat{\sigma}^2/\sigma^2}$ was employed.

The above expression in (1.6.8) can be transformed in terms of this statistic to give

$$E(q|A_1)P(A_1) = 2\theta \int_{t>\frac{m\lambda}{n-k}} f_t(t;m+4,n-k,\theta)dt \qquad (1.6.9)$$

$$+ m \int_{t>\frac{m\lambda}{n-k}} f_t(t;m+2,n-k,\theta)dt \quad .$$

The second integral is the function $r(\theta,\lambda)$ defined in section 1.5 to be $r(\theta,\lambda) = P\{\frac{m+2}{n-k} F'(m+2,n-k;\theta) \geq \frac{m}{n-k}\lambda\}$. Similarly, define $s(\theta,\lambda) = P\{\frac{m+4}{n-k} F'(m+4,n-k;\theta) \geq \frac{m}{n-k}\lambda\}$. From (1.6.9), and recalling that $q = \frac{Q}{\sigma^2}$, it follows that

$$E(Q|A_1)P(A_1) = \sigma^2\{2\theta\ s(\theta,\lambda) + m\ r(\theta,\lambda)\} \quad . \qquad (1.6.10)$$

As $\frac{Q}{\sigma^2} \sim \chi'^2(m;\theta)$, it follows from (1.6.7) that $E(\frac{Q}{\sigma^2}) = m + 2\theta$. Thus

$$E(Q|A_o)P(A_o) = \sigma^2\{m + 2\theta - mr(\theta,\lambda) - 2\theta s(\theta,\lambda)\}. \qquad (1.6.11)$$

To complete the task of this section, there are two remaining terms of (1.6.5) which need to be evaluated so that $M(\theta,\lambda)$ may be found.

Consider the third term on the right of (1.6.5); viz.

$$E[(H\beta-h)'[HS^{-1}H']^{-1}(Hb-h)|A_o]P(A_o)$$

H, β, h and S all consist of fixed constants and from (1.5.12) and (1.5.17)

$$E(b|A_o)P(A_o) = \beta(1-g(\theta,\lambda)) - S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)[r(\theta,\lambda)-g(\theta,\lambda)] \ .$$

Thus,

$$E(Hb|A_o)P(A_o) = H\beta(1-g(\theta,\lambda)) - (H\beta-h)[r(\theta,\lambda) - g(\theta,\lambda)] \quad (1.6.12)$$

$$= H\beta(1-r(\theta,\lambda)) + h[r(\theta,\lambda) - g(\theta,\lambda)]$$

and, clearly,

$$E(h|A_o)P(A_o) = h(1-g(\theta,\lambda)) \qquad (1.6.13)$$

and hence

$$E((H\beta-h)'[HS^{-1}H']^{-1}(Hb-h)|A_o)P(A_o) \qquad (1.6.14)$$

$$= (H\beta-h)'[HS^{-1}H']^{-1}(H\beta-h)(1-r(\theta,\lambda))$$

$$= 2\sigma^2\,\theta(1-r(\theta,\lambda)) \text{ from the definition of } \theta \text{ in } (1.3.3).$$

The remaining term to be evaluated in (1.6.5) is merely the transpose of this, which gives the same value as found in (1.6.14) for it is a scalar quantity. The quadratic risk function for y^* can now be written

$$M(\theta,\lambda) = \sigma^2\{k + 4\theta(1-r(\theta,\lambda)) - [m+2\theta-mr(\theta,\lambda)-2\theta s(\theta,\lambda)]\} \qquad (1.6.15)$$

$$= \sigma^2\{k-m + mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} .$$

1.7 Invariance under Orthogonal Transformations

In the discussion so far, no restrictions of orthogonality have been placed on the X's. Consider, now, an orthogonal transformation of the X's, defined by

$$Z = XC \text{ where } C \text{ is a } k\text{x}k \text{ orthogonal matrix.}$$

That is, $CC' = I$ and $C'X'XC = Z'Z$ where $Z'Z$ is a diagonal matrix. Suppose, now, that the parameter space and the H matrix are also transformed as follows:

$$\alpha = C'\beta \quad \text{and} \quad H_* = HC .$$

The non-centrality of (1.3.6) becomes

$$\theta = \frac{(H\beta-h)'[HS^{-1}H']^{-1}(H\beta-h)}{2\sigma^2} \tag{1.7.1}$$

$$= \frac{(HCC'\beta-h)'[HCC'S^{-1}CC'H]^{-1}(HCC'\beta-h)}{2\sigma^2} .$$

Now, $CC'=I$ implies that $C^{-1} = C'$, and

$$C'S^{-1}C = C^{-1}S^{-1}C'^{-1} = (C'SC)^{-1} = (C'X'XC)^{-1} \tag{1.7.2}$$

$$= (Z'Z)^{-1} .$$

θ in (1.7.1) can be written

$$\theta = \frac{(H_*\alpha-h)'[H_*(Z'Z)^{-1}H_*']^{-1}(H_*\alpha-h)}{2\sigma^2} . \tag{1.7.3}$$

These transformations obviously leave θ invariant. As the quadratic risk function, $M(\theta,\lambda)$ of (1.6.15), is expressed in terms of θ, it is invariant under this set of orthogonal transformations. Consider, now, the bias function of (1.5.18).

$$\text{bias } y^* = XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(r(\theta,\lambda)-1)$$

$$= XCC'S^{-1}CC'H'[HCC'S^{-1}CC'H]^{-1}(HCC'\beta-h)(r(\theta,\lambda)-1) .$$

In the light of (1.7.2) above, the bias function becomes:

$$\text{bias } y^* = Z(Z'Z)^{-1}H_*'[H_*(Z'Z)^{-1}H_*']^{-1}(H_*\beta-h)(r(\theta,\lambda)-1) .$$

Clearly, the bias function as well as the quadratic risk function remains invariant under the above set of orthogonal transformations.

2. FURTHER COMMENTS ON THE RESULTS OF CHAPTER 1

2.1 Comparison with Larson and Bancroft's Results

Larson and Bancroft [1963] addressed themselves to the problem of finding an estimator for the expected value of y, that is y^*, for a true population model, $y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + e$. The x's were divided into two groups. The first group $\{x_1, x_2, \ldots, x_m; m<k\}$ consisted of those variables which the experimenter felt were necessary for accurate prediction. A preliminary F test was used to decide on whether to include the other group $\{x_{m+1}, x_{m+2}, \ldots, x_k\}$ in the prediction model.

In other words, an F test was used to test the hypothesis $H_o: \beta_{m+1} = \beta_{m+2} = \ldots = \beta_k = 0$ using as the test statistic,

$$F_o = \frac{b_{m+1}^2 + b_{m+2}^2 + \ldots + b_k^2}{(k-m)V}$$ where V is the

experimental error mean square.

Thus, the predicted value of y, y^*, can be written:

$$y^* = \begin{cases} b_1x_1 + b_2x_2 + \ldots + b_kx_k & \text{if } F_o \geq \lambda \\ b_1x_1 + b_2x_2 + \ldots + b_mx_m & \text{if } F_o < \lambda \end{cases}$$

where λ is an appropriate critical value of Snedecor's F test, (say) for a type I error of 5 percent. They obtained the result that

$$\text{bias of } y^* = [1-h(\theta)] \sum_{i=m+1}^{k} \beta_i x_i \qquad (2.1.1)$$

where $h(\theta)$ is the $r(\theta,\lambda)$ used in this paper.

To obtain this result, they assumed that the x's were orthogonal, and indeed were so scaled that

$$b_i \sim N(\beta_i, \sigma^2) \quad \text{for } i = m+1, m+2, \ldots, k \quad .$$

Under their assumptions and using zero restrictions as they did, their formulation is equivalent to taking h to be the zero vector, and $H = (0_{k-m,m}\ , I_{k-m})$ where I_{k-m} is an identity matrix of rank (k-m), and $0_{k-m,m}$ is a (k-m)xm matrix of zeros. Also, their X'X matrix would take the form $\begin{pmatrix} D_{m,m} & 0_{m,k-m} \\ 0_{k-m,m} & I_{k-m} \end{pmatrix}$ where $D_{m,m}$ is a diagonal mxm matrix.

$HS^{-1}H'$ would be $(0_{k-m,m}\ I_{k-m}) \begin{pmatrix} D^{-1}_{m,m} & 0_{m,k-m} \\ 0_{k-m,m} & I_{k-m} \end{pmatrix} \begin{pmatrix} 0_{m,k-m} \\ I_{k-m} \end{pmatrix}$

which simplifies to I_{k-m}.

The bias function as given in (1.5.18) was

$$\text{bias } y^* = XS^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(r(\theta,\lambda)-1) \quad . \tag{2.1.2}$$

Under Larson and Bancroft's assumptions, the trace of the expression in (2.1.2) reduces to the expression in (2.1.1). Furthermore, it is of interest to note that the general quadratic risk function found in (1.6.15) to be:

$$M(\theta,\lambda) = \sigma^2\{k-m+mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda)+s(\theta,\lambda)]\} \tag{2.1.3}$$

reduces with their simpler model to their results that:

Mean square error of y^*

$$= \sigma^2\{\frac{1}{n} + \sum_{i=1}^{m} x_i^2 + h(\theta) \sum_{i=m+1}^{k} x_i^2 + [r(\theta)-2h(\theta)+1](\sum_{i=m+1}^{k} \frac{\beta_i}{\sigma} x_i)^2\} \tag{2.1.4}$$

The $r(\theta)$ in (2.1.4) is the $s(\theta,\lambda)$ defined in section 1.6, and the term "$\frac{1}{n}$" enters into their expression because they included a constant term, "β_o" in their model. Another apparent discrepancy between (2.1.3) and (2.1.4) is that m independent restrictions are used in this paper whereas Larson and Bancroft used k-m, and because of the scaling mentioned above for the b_i, $\sum_{i=m+1}^{k} x_i^2 = (k-m)$. As the x_i are orthogonal, then $(\sum_{i=m+1}^{k} \frac{\beta_i x_i}{\sigma})^2$ reduces to 2θ. By a suitable scaling, $\sum_{i=1}^{m} x_i^2$ is seen to be equal to m, which is analogous to the term "k-m" in (2.1.3).

Although Larson and Bancroft developed the expression (2.1.1) and (2.1.4) by assuming orthogonality of the x's, they included a proof by David that the bias function would be unaltered by non-orthogonality. The proof given relies on a transformation of the sample space of x's, and of the parameter space of b's. In section 1.8, it was shown that under such transformations not only the bias of y^* but also the quadratic risk function remains invariant, and that this is so for general linear restrictions. However, they appear to have overlooked the point that for invariance the H matrix must also be transformed, which raises a problem for the case of zero restrictions. If Larson and Bancroft are working with the transformed model of Z, α and H_*, as the appendix seems to imply, then H_* would equal $(0_{k-m,m}, I_{k-m})$. But in the untransformed model, $H \neq (0_{k-m,m}, I_{k-m})$, so that the original restrictions would not be zero restrictions but messy restrictions of the form of linear combinations of the β's being zero.

2.2 Comparison with Wallace's Results

T. D. Wallace[4] introduced the term "weak mean squared error" and defined it as:

> "The restricted estimator, $\hat{\beta}$, is better in weak mean squared error if and only if $E(\hat{\beta}-\beta)'S(\hat{\beta}-\beta) \leq E(b-\beta)'S(b-\beta)$."

For the set of restrictions introduced in section 1.2, he showed that $\hat{\beta}$ is better in weak mean squared error if and only if the non-centrality parameter $\theta \leq {}^{m}/_{2}$. This provides a method of checking the expression from (1.6.15), that

$$M(\theta,\lambda) = \sigma^2[k-m+mr(\theta,\lambda) + 2\theta(1-2r(\theta,\lambda) + s(\theta,\lambda))] \quad .$$

b, the o.l.s. estimate of β, would always be chosen if the critical value of the u-statistic, λ, is zero. From the definitions of $r(\theta,\lambda)$ and $s(\theta,\lambda)$, they would both become 1 if $\lambda = 0$.

That is,

$$E(b-\beta)'S(b-\beta) = M(\theta,0) = k\sigma^2 \quad . \tag{2.2.1}$$

On the other hand, as $\lambda \rightarrow \infty$, $r(\theta,\lambda)$ and $s(\theta,\lambda)$ both tend to zero, and $\hat{\beta}$ would be chosen with probability 1. That is

$$E(\hat{\beta}-\beta)'S(\hat{\beta}-\beta) = M(\theta,\infty) = \sigma^2[k-m+2\theta] \quad . \tag{2.2.2}$$

Thus, $\hat{\beta}$ is better in weak mean squared error if and only if

---

[4] Unpublished paper. See footnote 1.

$M(\theta,\infty) \leq M(\theta,0)$ which implies that

$$\sigma^2[k-m+2\theta] \leq \sigma^2 k$$

which in turn implies that $\theta \leq m/2$.

This provides a useful check on the expression given for $M(\theta,\lambda)$.

2.3 Comparison with Sawa and Hiromatsu's Results

Sawa and Hiromatsu [1971] set up a model similar to Larson and Bancroft [1963], with orthogonal x's and q zero restrictions. Evidently aiming to focus attention on the β vector itself rather than the predicted value of y, they considered the following estimator for the linear combination of β's, $\underset{\sim}{C}'\underset{\sim}{\beta}$:

$$\phi_\lambda = \begin{cases} \underset{\sim}{C}_1' \underset{\sim}{b}_1 + \underset{\sim}{C}_2' \underset{\sim}{b}_2 & \text{if } F_o \geq \lambda \\ \underset{\sim}{C}_1' \underset{\sim}{b}_1 & \text{if } F_o < \lambda \end{cases}$$

where $F_o = \dfrac{\underset{\sim}{b}_2' \underset{\sim}{b}_2}{q\, s^2}$ and s^2 is the sample variance.

They obtained the result:

$$\text{Mean square error } \phi_\lambda = \sigma^2[\underset{\sim}{C}_1' \underset{\sim}{C}_1 + h(\theta,\lambda)\, \underset{\sim}{C}_2' \underset{\sim}{C}_2] + (\underset{\sim}{\beta}_2' \underset{\sim}{C}_2)^2[1-2h(\theta,\lambda) + s(\theta,\lambda)] \qquad (2.3.1)$$

and $h(\theta,\lambda)$ is analogous to the $r(\theta,\lambda)$ of section 1.5. This suffers from the disadvantage that this expression depends on the vectors $\underset{\sim}{C}_1$ and $\underset{\sim}{C}_2$. To overcome this, they defined a risk function:

$$R(\theta,\lambda) = \sup_{\underset{\sim}{C}_2} [E(\hat{\phi}_\lambda - \phi)^2 - \sigma^2 \underset{\sim}{C}_1' \underset{\sim}{C}_1] \Big/ (\sigma^2 \underset{\sim}{C}_2' \underset{\sim}{C}_2) \qquad (2.3.2)$$

which gives

$$R(\theta,\lambda) = h(\theta,\lambda) + \theta[s(\theta,\lambda) - 2h(\theta,\lambda)+1] \qquad (2.3.3)$$

where their θ is twice the non-centrality used in Chapter 1. Unfortunately, they are rather careless in their definitions as on page 6 they define $h(\theta,\lambda) \equiv P\{F'(q+2,n-k;\theta) \geq \lambda\}$

and $\qquad s(\theta,\lambda) \equiv P\{F'(q+4,n-k;\theta) \geq \lambda\}$

whereas they should have included constant factors $\frac{m}{n-k}$, $\frac{m+2}{n-k}$, $\frac{m+4}{n-k}$, as was shown in section 1.6. Furthermore, they are not consistent as they give another incorrect version in a footnote on page 14, stating:

$$h(\theta,\lambda) = e^{-\theta/2} \sum_{j=o}^{\infty} \frac{(\theta/2)^j}{j!} P\{\chi^2_{q+2+2j} \geq \lambda\chi^2_{n-k}\} \ . \qquad (2.3.4)$$

This should be, using their definition of θ,

$$h(\theta,\lambda) = e^{-\theta/2} \sum_{j=o}^{\infty} \frac{(\theta/2)^j}{j!} P\{\chi^2_{q+2+2j} \geq \frac{q}{n-k}\lambda\chi^2_{n-k}\} \ . \qquad (2.3.5)$$

Using (2.3.4) they omitted a factor of 1/2 on the right-hand side when they obtained:

$$\frac{\partial}{\partial\theta} h(\theta,\lambda) = s(\theta,\lambda) - h(\theta,\lambda) \quad . \qquad (2.3.6)$$

Fortunately, these omissions have not invalidated the risk function they obtained, which is quoted above in (2.3.2).

It is interesting to notice the similarity of form of their expression in (2.3.1) and the quadratic risk function of (1.6.15). Further comments on their risk function will be made in section 4.6.

3. MINIMAX REGRET FUNCTION FOR y^*

3.1 A Minimax Regret Function Defined

It has been shown in section 1.7 that the quadratic risk function, $M(\theta,\lambda)$, based on an estimator for the predicted value of y, is invariant under groups of orthogonal transformations of the sample, parameter and restriction spaces. It is appealing, then, to define a minimax regret function based on this quadratic risk function in an attempt to find an optimal critical value, λ^*, of the prior test of estimation.

When values of the quadratic risk function of y^*, $M(\theta,\lambda)$, are computed for a given value of $\lambda=\lambda_o$ in the open interval $(0,+\infty)$, it is found that the graph of $M(\theta,\lambda_o)$ follows the general shape of the curve labelled $\lambda=\lambda_o$ in Figure 3.1. The quadratic risk functions for Xb and $X\hat{\beta}$ are represented by the lines labelled $\lambda=0$ and $\lambda=\infty$, respectively.

The curve of $M(\theta,\lambda_o)$ for λ_o in $(0,+\infty)$ has the following characteristics: When the non-centrality parameter, θ, is zero, $M(\theta,\lambda_o)$ is between q.r.f. $(X\hat{\beta})$, the quadratic risk function for $X\hat{\beta}$, and q.r.f. (Xb). As θ increases, $M(\theta,\lambda_o)$ increases and its graph cuts that of q.r.f. (Xb) at a point between $\theta = \frac{m}{4}$ and $\theta = \frac{m}{2}$. $M(\theta,\lambda_o)$ then reaches a maximum at a point $\theta=\theta_U$ where $\theta_U \geq \frac{m}{2}$. It would appear that the graph of $M(\theta,\lambda_o)$ is unimodal, but this is not essential to the definition of the minimax regret condition given below.

If θ were known, then the best estimator of y^*, as far as the quadratic risk function is concerned, would be $X\hat{\beta}$ for $\theta < \frac{m}{2}$, and Xb for $\theta \geq \frac{m}{2}$.

Consider $\inf_\lambda M(\theta,\lambda)$, that is the infimum (which in this context will equal the minimum) of $M(\theta,\lambda)$ over all values of λ. Clearly,

$$\inf_{\lambda} M(\theta,\lambda) = \begin{cases} M(\theta,+\infty) \text{ or q.r.f. } (X\hat{\beta}), \text{ for } \theta < \frac{m}{2} \\ M(\theta,0), \text{ or q.r.f. } (Xb), \text{ for } \theta \geq \frac{m}{2} \end{cases} \quad . \tag{3.1.1}$$

The minimax regret function, $REG(\theta,\lambda)$, is then defined as

$$REG(\theta,\lambda) = M(\theta,\lambda) - \inf_{\lambda} M(\theta,\lambda) \quad . \tag{3.1.2}$$

For $\theta < \frac{m}{2}$ and small values of m and n-k, it is found, empirically, that $REG(\theta,\lambda)$ takes on a maximum at $\theta=0$. For larger values of m and n-k, there is a θ_L at which the maximum occurs. In Figure 3.1, this value of $REG(\theta,\lambda)$ is labelled δ_L. For $\theta \geq \frac{m}{2}$, $REG(\theta,\lambda)$ takes on a maximum value at θ_U and this value of $REG(\theta,\lambda)$ is labelled δ_U in Figure 3.1. Heuristically, these two distances can be thought of as the maximum additional penalty for choosing $X\beta^*$ instead of the optimal estimators, $X\hat{\beta}$ and Xb, at the points $\theta=\theta_L$ and $\theta=\theta_U$, respectively.

The minimax regret procedure is to seek the $\lambda=\lambda^*$ which makes both δ_L and δ_U as small as possible. It is found, however, that as the critical value, λ, increases the distance labelled δ_L decreases but δ_U increases. This can be seen in Figure 3.1 by the relative positions of the graphs labelled $\lambda=1$ and $\lambda=4$.

Thus, to minimize the minimax regret function, $REG(\theta,\lambda)$, over all values of θ and λ, the procedure will be to seek a $\lambda=\lambda^*$ such that

$$REG(\theta_L,\lambda^*) = REG(\theta_U,\lambda^*) \quad , \tag{3.1.3}$$

or, in other words,

$$\sup_{\theta<\frac{m}{2}} REG(\theta,\lambda^*) = \sup_{\theta\geq\frac{m}{2}} REG(\theta,\lambda^*) \quad . \tag{3.1.4}$$

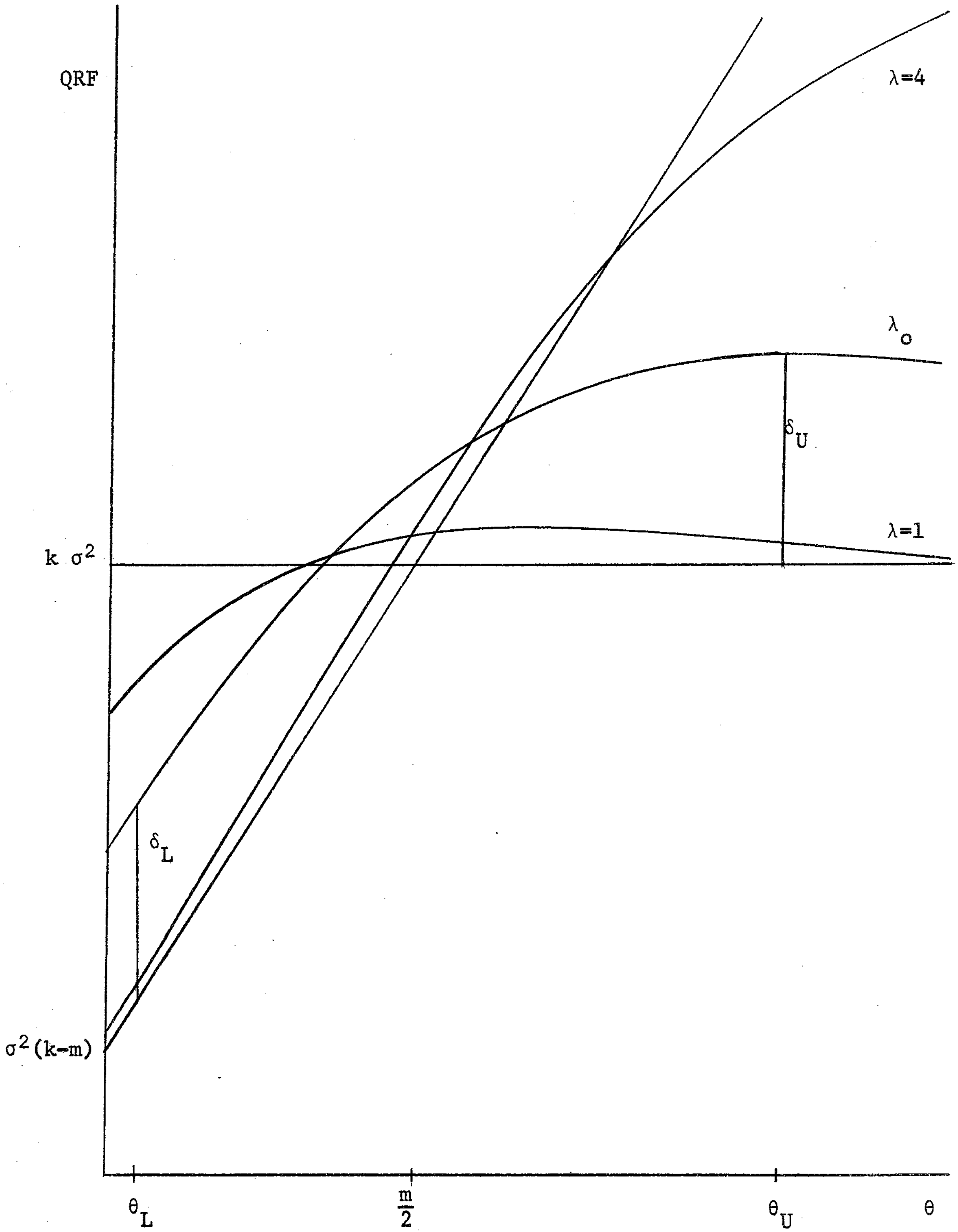


Figure 3.1. Quadratic risk functions for different λ

In terms of the quadratic risk function of y^*, the aim is to find λ^* such that

$$M(\theta_L,\lambda^*) - M(\theta_L,+\infty) = M(\theta_U,\lambda^*) - M(\theta_U,0) \quad . \tag{3.1.5}$$

If the expression for $M(\theta,\lambda)$, as found in (1.6.15), is substituted in (3.1.5), the minimax regret condition implies the following relationship:

$$\begin{aligned} & m\, r(\theta_L,\lambda^*) + 2\theta_L[-2r(\theta_L,\lambda^*) + s(\theta_L,\lambda^*)] \\ = \;& m\, r(\theta_U,\lambda^*) - m + 2\theta_U[1-2r(\theta_U,\lambda^*) + s(\theta_U,\lambda^*)] \end{aligned} \tag{3.1.6}$$

where θ_L, θ_U maximize the quantities on the left and right, respectively.

3.2 Some Properties of the Functions $M(\theta,\lambda)$, $r(\theta,\lambda)$ and $s(\theta,\lambda)$

In section 3.1, some general properties were noted of the graph of $M(\theta,\lambda)$. These properties will be dealt with in this section more formally by a series of lemmas.

<u>Lemma 3.2.1</u>

$M(\theta,0)$ = q.r.f. (Xb) is a constant function of θ, and $M(\theta,+\infty)$ = q.r.f. $(X\hat{\beta})$ is a linear function of θ.

<u>Proof</u>

Recall that

$$M(\theta,\lambda) = \sigma^2\{k-m+mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \quad . \tag{3.2.1}$$

From the definitions of $r(\theta,\lambda)$, $s(\theta,\lambda)$ given in section 1.6, it is clear that $r(\theta, 0)$ and $s(\theta, 0)$ are both unity, and $r(\theta,\infty)$, $s(\theta,\infty)$ are both zero. Substituting these values in (3.2.1) gives the results:

$$M(\theta,0) = k\sigma^2 \tag{3.2.2}$$

and

$$M(\theta,\infty) = \sigma^2\{k-m + 2\theta\} \quad . \tag{3.2.3}$$

Clearly, the quadratic risk functions for Xb and $X\hat{\beta}$ are represented by the lines labelled $\lambda=0$ and $\lambda=\infty$, respectively. $M(\theta,\lambda)$ inherits many of its characteristics from the functions $r(\theta,\lambda)$ and $s(\theta,\lambda)$. For this reason, attention will be focused on these functions in the next two lemmas. Now $r(\theta,\lambda)$ was defined after (1.5.18) in terms of the ratio of two sums of squares, that is

$$r(\theta,\lambda) = P\{\frac{m+2}{n-k} F'(m+2,n-k;\theta) \geq \frac{m\lambda}{n-k}\} \quad . \tag{3.2.4}$$

Equivalently,

$$r(\theta,\lambda) = P\{\chi'^2 (m+2;\theta) \geq \frac{m\lambda}{n-k} \chi^2(n-k)\} \quad . \tag{3.2.5}$$

By a well-known transformation, for example see Wilks [1943, p. 187], $r(\theta,\lambda)$ can be expressed in terms of the incomplete Beta function. If $X \sim F(2p,2q)$, then $Y = \frac{pX}{pX+q}$ has the Beta distribution with parameters p and q, and $P\{Y<y\} = I_y(p,q)$ where $I_y(p,q)$ is the incomplete Beta function ratio and

$$I_y(p,q) = \frac{1}{B(p,q)} \int_0^y t^{p-1} (1-t)^{q-1}\, dt \tag{3.2.6}$$

and $B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ can be termed the complete Beta function. With this notation,

$$r(\theta,\lambda) = 1 - \sum_{j=0}^{\infty} \frac{e^{-\theta}\,\theta^j}{j!}\, I_y(\frac{m+2}{2} + j,\ \frac{n-k}{2}) \tag{3.2.7}$$

and $y = \frac{m\lambda}{m\lambda+n-k}$.

$s(\theta,\lambda)$ will have the same form as (3.2.7) with (m+4) replacing the factor (m+2). A third function $w(\theta,\lambda)$ can be defined with (m+6) replacing (m+2). Perhaps, more succinct notation would be the following:

$$r(\theta,\lambda) = I'(\tfrac{m}{2}),\ s(\theta,\lambda) = I'(\tfrac{m}{2}+1),\ w(\theta,\lambda) = I'(\tfrac{m}{2}+2) \qquad (3.2.8)$$

$$s(\theta,\lambda) - r(\theta,\lambda) = \Delta I'(\tfrac{m}{2})$$
$$w(\theta,\lambda) - 2s(\theta,\lambda) + r(\theta,\lambda) = \Delta^2 I'(\tfrac{m}{2}) \quad . \qquad (3.2.9)$$

Lemma 3.2.2

For θ and λ in the open interval $(0, +\infty)$,

$$r(\theta,\lambda) < s(\theta,\lambda) < w(\theta,\lambda) \quad .$$

In other words, $r(\theta,\lambda)$ is an increasing function of the numerator degrees of freedom. Equivalently, $\Delta I'(p)>0$ for all integer p.

Proof

The result can also be obtained from the difference

$$s(\theta,\lambda) - r(\theta,\lambda) = \sum_{j=o}^{\infty} \frac{e^{-\theta}\,\theta^j}{j!}\left[I_y(\tfrac{m+2}{2}+j,\ \tfrac{n-k}{2}) - I_y(\tfrac{m+4}{2}+j,\ \tfrac{n-k}{2})\right] . \qquad (3.2.10)$$

Jordan [1962, p. 84] gives the following useful relationship:

$$I_y(p,q) = \frac{\Gamma(p+q)}{\Gamma(p+1)\Gamma(q)}\, y^p(1-y)^q + I_y(p+1,\ q) \quad . \qquad (3.2.11)$$

In other words, $I_y(p+1,\ q) - I_y(p,q) = \frac{\Gamma(p+q)}{\Gamma(p+1)\Gamma(q)}\, y^p(1-y)^q$.

The expression in (3.2.10) can then be written as

$$s(\theta,\lambda) - r(\theta,\lambda) = \sum_{j=o}^{\infty} \frac{e^{-\theta}\theta^{j}}{j!} \frac{\Gamma(\frac{m+2+2j+n-k}{2})}{\Gamma(\frac{m+4+2j}{2})\Gamma(\frac{n-k}{2})} y^{\frac{m+2+2j}{2}} (1-y)^{\frac{n-k}{2}} > 0 \text{ as } y$$

is in the interval (0, 1). Consequently,

$$s(\theta,\lambda) > r(\theta,\lambda) \tag{3.2.12}$$

$w(\theta,\lambda) > s(\theta,\lambda)$ follows from a similar argument.

Lemma 3.2.3

For θ,λ in the open interval $(0, +\infty)$, $r(\theta,\lambda)$, $s(\theta,\lambda)$ and $w(\theta,\lambda)$ are decreasing functions of λ, but increasing functions of θ.

Proof

It is clear from the definition of $r(\theta,\lambda)$ in (3.2.5) that as λ increases then $r(\theta,\lambda)$ decreases. It was shown in (1.5.19) that

$$\frac{\partial g(\theta,\lambda)}{\partial\theta} = - g(\theta,\lambda) + r(\theta,\lambda) .$$

In a similar way, or directly from (3.2.7),

$$\frac{\partial r(\theta,\lambda)}{\partial\theta} = - r(\theta,\lambda) + s(\theta,\lambda) \tag{3.2.13}$$

and this expression is positive from Lemma 3.2.2. The lemma is now proved for $r(\theta,\lambda)$, and analogously, the results follow for $s(\theta,\lambda)$ and $w(\theta,\lambda)$.

Lemma 3.2.4

When $\theta=0$, and λ in $(0, +\infty)$, $M(\theta,\lambda)$ lies between the quadratic risk function for $X\hat{\beta}$ and the quadratic risk function for Xb. That is,

$$M(0, +\infty) < M(0,\lambda) < M(0,0) .$$

Proof

Now $r(0,\lambda)$ and $s(0,\lambda)$ are both unity when $\lambda=0$ and both tend to zero as λ tends to infinity. Thus,

$$M(0,\lambda) = \sigma^2\{k-m+mr(0,\lambda)\} \tag{3.2.14}$$

and clearly this lies between $M(0, +\infty) = \sigma^2\{k-m\}$ and $M(0,0) = k\sigma^2$.

Lemma 3.2.5

The graph $M(\theta,\lambda)$, the quadratic risk function of $X\beta^*$, cuts that of Xb at $\theta=\theta_c$ where θ_c is in the open interval $(\frac{m}{4}, \frac{m}{2})$, provided that λ is in $(0, +\infty)$.

Proof

From (3.2.1) and (3.2.2), θ_c is defined by

$$k\sigma^2 = \sigma^2\{k-m+mr(\theta_c,\lambda) + 2\theta_c[1-2r(\theta_c,\lambda) + s(\theta_c,\lambda)]\} \tag{3.2.15}$$

which simplifies to

$$\theta_c = \frac{m}{2}\left[\frac{1-r(\theta_c,\lambda)}{1-2r(\theta_c,\lambda) + s(\theta_c,\lambda)}\right] . \tag{3.2.16}$$

The lemma will be proved if it can be shown that

$$\frac{1}{2} < \frac{1 - r(\theta_c,\lambda)}{1-2r(\theta_c,\lambda) + s(\theta_c,\lambda)} < 1$$

or

$$\frac{1}{2} < \frac{1 - r(\theta_c,\lambda)}{1-r(\theta_c,\lambda) + s(\theta_c,\lambda) - r(\theta_c,\lambda)} < 1 . \tag{3.2.17}$$

The right-hand inequality in (3.2.17) follows from the fact that $s(\theta_c,\lambda) - r(\theta_c,\lambda) > 0$ as shown in (3.2.12). As $s(\theta_c,\lambda)$ is a probability, it is less than 1 for λ in $(0, +\infty)$, so that the denominator,

$$1 - r(\theta_c,\lambda) + s(\theta_c,\lambda) - r(\theta_c,\lambda) < 2[1 - r(\theta_c,\lambda)]$$

which proves the left-hand inequality, and completes the proof of the lemma.

Lemma 3.2.6

The infimum of $M(\theta,\lambda)$ is $M(\theta, +\infty)$, the q.r.f. $(X\hat{\beta})$, when $\theta < \frac{m}{2}$, and is $M(\theta,0)$, the q.r.f. (Xb) when $\theta \geq \frac{m}{2}$.

Proof

Now $s(\theta,\lambda) - r(\theta,\lambda) > 0$ from Lemma 3.2.2. For $\theta < \frac{m}{2}$,

$$\begin{aligned} M(\theta,\lambda) &= \sigma^2\{k-m+mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \\ &> \sigma^2\{k-m+2\theta+mr(\theta,\lambda) - 2\theta r(\theta,\lambda)]\} \\ &> \sigma^2\{k-m+2\theta\} = M(\theta, +\infty) . \end{aligned}$$

Also, $1-2r(\theta,\lambda) + s(\theta,\lambda) > 0$, hence for $\theta \geq \frac{m}{2}$

$$\begin{aligned} M(\theta,\lambda) &\geq \sigma^2\{k-m+mr(\theta,\lambda) + m[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \\ &= \sigma^2\{k + m[s(\theta,\lambda) - r(\theta,\lambda)]\} \\ &> \sigma^2 k = M(\theta, 0) . \end{aligned}$$

Lemma 3.2.7

As θ tends to infinity, $M(\theta,\lambda)$ approaches $M(\theta,0)$, the quadratic risk function for Xb.

Proof

$$M(\theta,\lambda) = \sigma^2\{k-m+mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} .$$

Gun [1965, p. 54] has shown that, with the present notation,

$$\theta^a[1-r(\theta,\lambda)] \to 0 \text{ as } \theta \to +\infty, \text{ and } a > 0 \quad . \qquad (3.2.18)$$

Thus, $2\theta[2\{1-r(\theta,\lambda)\} - \{1-s(\theta,\lambda)\}] \to 0$ as $\theta \to +\infty$. As $r(\theta,\lambda)$ is an increasing function of θ, then $\lim_{\theta\to+\infty} r(\theta,\lambda) = 1$.
Hence, the result that

$$M(\theta,\lambda) \to k\sigma^2 \quad \text{as} \quad \theta \to +\infty,$$

As $r(\theta,\lambda)$ and $s(\theta,\lambda)$ are continuous functions of both θ and λ, it is clear that the regret function of (3.1.2), $REG(\theta,\lambda) = M(\theta,\lambda) - \inf_{\lambda} M(\theta,\lambda)$ will attain a maximum value for θ in the interval $[0, \frac{m}{2}]$, and $[\frac{m}{2}, +\infty]$ as $REG(\theta,\lambda)$ decreases to zero as θ goes to infinity.

4. MINIMAX REGRET FUNCTION FOR THE BETA VECTOR

4.1 Introductory Comments

In Chapter 1, the estimator, $X\beta^*$, of the predicted value of y was studied and its bias and its quadratic risk function were obtained as neat, mathematical expressions. On the basis of the latter, a minimax regret function was defined in Chapter 3 to seek an optimum critical point for the preliminary test of estimation.

In this chapter, the bias and the mean square error of β^* will be found by similar methods to those used in Chapter 1. In contrast to the bias and quadratic risk function of $X\beta^*$, the bias and mean square error of β^* will not be invariant under orthogonal transformations. It will not be possible to find an optimal critical value independent of the design matrix, $S = X'X$, or the restriction matrix, H. In a given experiment when S and H are specified, it will be possible to find an optimal critical value, λ_o, by setting up a similar minimax condition as used in Chapter 3.

Differentiation of matrices and vectors will again be greatly involved, and it will be helpful to recall the expression found in (1.5.7):

$$\frac{\partial}{\partial\beta}[(b-\beta)'S(b-\beta)] = -2S(b-\beta) .$$

Differentiating with respect to β' would give the transpose of this, namely,

$$\frac{\partial}{\partial\beta'}[(b-\beta)'S(b-\beta)] = -2(b-\beta)'S , \qquad (4.1.1)$$

as S is symmetric.

This expression can be differentiated a second time, as shown in Theil [1971, p. 31], giving the result

$$\frac{\partial^2}{\partial\beta\partial\beta'}[(b-\beta)'S(b-\beta)] = 2S \quad . \tag{4.1.2}$$

4.2 Bias of the Beta Vector

The model, restrictions, and test statistic are the same as in Chapter 1. The estimator of the β vector is given by

$$\beta^* = \begin{cases} b & \text{if } u \geq \lambda \quad \text{on the set } A_1 \quad , \\ \hat{\beta} & \text{if } u < \lambda \quad \text{on the set of } A_0 \quad . \end{cases}$$

The bias β^* can be found in a similar way to the bias y^* found in (1.5.20), thus

$$\text{bias } \beta^* = S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(r(\theta,\lambda)-1) \quad . \tag{4.2.1}$$

As a partial check on this bias function, it is obvious that it takes the value zero if the restrictions are exact, that is if $H\beta=h$, or if λ takes the value zero. This latter case is equivalent to always choosing the o.l.s. estimator of β, namely b, which is, of course, unbiased.

4.3 The Mean Square Error of the Estimated Beta Vector

In this section, the mean square error of β^* will be evaluated, that is $E(\beta^*-\beta)(\beta^*-\beta)'$. In section 1.6, it may be recalled, it was the trace of the mean square error of y^*, that is $E(X\beta^*-X\beta)'(X\beta^*-X\beta)$, which was the more easily obtained.

It is convenient to consider the transpose of each of the expressions in (1.5.8) through (1.5.10).

$$\frac{\partial P(A_1)}{\partial \beta'} = \iint_{A_1} K\, f_2(V)\, \frac{(b-\beta)'S}{\sigma^2} \exp\left[-\frac{(b-\beta)'S(b-\beta)}{2\sigma^2}\right] dbdV \qquad (4.3.1)$$

and

$$\frac{\partial P(A_1)}{\partial \beta'} = \frac{\partial g(\theta,\lambda)}{\partial \theta}\, \frac{(H\beta-h)'[HS^{-1}H']^{-1}H}{\sigma^2} \; . \qquad (4.3.2)$$

The mean square error of β^* will contain an expression $E[(b-\beta)(b-\beta)'|A_o]P(A_o)$ and to evaluate this, it is necessary to differentiate (4.3.2) a second time to obtain:

$$\frac{\partial^2 P(A_1)}{\partial \beta \partial \beta'} = \iint_{A_1} K\, f_2(V)\, \{-\frac{S}{\sigma^2} + \frac{S(b-\beta)(b-\beta)'S}{\sigma^4}\} \exp\left[-\frac{(b-\beta)'S(b-\beta)}{2\sigma^2}\right] dbdV. \qquad (4.3.3)$$

On the other hand, from (4.3.2), the following can be obtained:

$$\frac{\partial^2 P(A_1)}{\partial \beta \partial \beta'} = g'(\theta,\lambda)\, \frac{H'[HS^{-1}H']^{-1}H}{\sigma^2} + g''(\theta,\lambda)\, \frac{H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}H}{\sigma^4} \; . \qquad (4.3.4)$$

Clearly, $\frac{\partial^2 P(A_1)}{\partial \beta \partial \beta'}$ is a kxk matrix whose $(i,j)^{th}$ term is $\frac{\partial^2 P(A_1)}{\partial \beta_i \partial \beta_j}$.

From (4.3.1) and (4.3.2), (or alternatively, see the derivation leading to (1.5.11))

$$E(b-\beta|A_1)P(A_1) = S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)\, g'(\theta,\lambda) \; . \qquad (4.3.5)$$

From (4.3.3) and (4.3.4),

$$E[-\frac{S}{\sigma^2} + \frac{S(b-\beta)(b-\beta)'S'}{\sigma^4} |A_1]P(A_1)$$

$$= g'(\theta,\lambda)\frac{H'[HS^{-1}H']^{-1}H}{\sigma^2} + g''(\theta,\lambda)\frac{H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}H}{\sigma^4} . \tag{4.3.6}$$

Now S is symmetric so that $S = S'$. Pre- and post-multiplying the expression in (4.3.6) by $S^{-1}\sigma^2$ and recalling from (1.5.2) that $g(\theta,\lambda)$ is $P(A_1)$, the following result is obtained:

$$E[(b-\beta)(b-\beta)'|A_1]P(A_1) = \sigma^2 S^{-1} g(\theta,\lambda)$$

$$+ \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}g'(\theta,\lambda) + S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'$$

$$[HS^{-1}H']^{-1}HS^{-1}g''(\theta,\lambda) . \tag{4.3.7}$$

Now

$$E[(b-\beta)(b-\beta)'|A_o]P(A_o) + E[(b-\beta)(b-\beta)'|A_1]P(A_1)$$

$$= \text{var } b = S^{-1}\sigma^2 .$$

Also $g'(\theta,\lambda) = -g(\theta,\lambda) + r(\theta,\lambda)$ by (1.5.19). Similarly, it can be shown that

$$r'(\theta,\lambda) = -r(\theta,\lambda) + s(\theta,\lambda) \quad \text{so that}$$

$$g''(\theta,\lambda) = g(\theta,\lambda) - 2r(\theta,\lambda) + s(\theta,\lambda) .$$

Substituting these values back in (4.3.7) leads to

$$E[(b-\beta)(b-\beta)'|A_o]P(A_o) = \sigma^2 S^{-1}[1-g(\theta,\lambda)]$$

$$- \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[-g(\theta,\lambda) + r(\theta,\lambda)]$$

$$- S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}[g(\theta,\lambda) - 2r(\theta,\lambda)$$

$$+ s(\theta,\lambda)] \quad . \qquad (4.3.8)$$

As $E[b-\beta|A_o]P(A_o) + E[b-\beta|A_1]P(A_1) = E(b-\beta) = 0$, then

$$E[b-\beta|A_o]P(A_o) = -S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)[-g(\theta,\lambda) + r(\theta,\lambda)] \quad . \qquad (4.3.9)$$

Denote the mean square error matrix of $\beta^* = E(\beta^*-\beta)(\beta^*-\beta)'$ by MSE (θ,λ). Then, as $\{A_o, A_1\}$ form a partition of the parameter space,

$$\begin{aligned} MSE(\theta,\lambda) &= E[(\beta^*-\beta)(\beta^*-\beta)'|A_1]P(A_1) + E[(\beta^*-\beta)(\beta^*-\beta)'|A_o]P(A_o) \\ &= E[(b-\beta)(b-\beta)'|A_1]P(A_1) + E[(\hat\beta-\beta)(\hat\beta-\beta)'|A_o]P(A_o) \quad . \end{aligned} \qquad (4.3.10)$$

As $\hat\beta = b - S^{-1}H'[HS^{-1}H']^{-1}(Hb-h)$, MSE (θ,λ) is found to be the sum of the following four terms:

(a) $E[(b-\beta)(b-\beta)']$, as this term appears in both sets A_o and A_1,

(b) $-S^{-1}H'[HS^{-1}H']^{-1}\ E[(Hb-h)(b-\beta)'|A_o]P(A_o)$,

(c) the transpose of (b), and

(d) $S^{-1}H'[HS^{-1}H']^{-1}\ E[(Hb-h)(Hb-h)'|A_o]P(A_o)[HS^{-1}H']^{-1}HS^{-1}$.

Note that in (b) and (d), use is made of the fact that S and H are matrices of known constants.

Now (a) $E[(b-\beta)(b-\beta)'] = \sigma^2 S^{-1}$ as $b \sim N(\beta, \sigma^2 S^{-1})$. (b) can be simplified using the identity

$$(Hb-h) = H(b-\beta) + (H\beta-h) , \qquad (4.3.11)$$

giving

$$(b) = -S^{-1}H'[HS^{-1}H']^{-1}E\{H(b-\beta)(b-\beta)' + (H\beta-h)(b-\beta)' | A_o\}P(A_o)$$

$$= -S^{-1}H'[HS^{-1}H']^{-1}H\, \eta_1$$

$$-S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)\, \eta_2 , \qquad (4.3.12)$$

where η_1 is the expression in (4.3.8) and η_2' is the transpose of the expression in (4.3.9).

On simplifying,

$$(b) = \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[r(\theta,\lambda) - 1]$$

$$+ S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}][s(\theta,\lambda) - r(\theta,\lambda)]. \qquad (4.3.13)$$

As this expression is symmetric, then (c) reduces to this same expression.

The expression (d) can be expanded into four terms as

$$(Hb-h)(Hb-h)' = H(b-\beta)(b-\beta)'H' + H(b-\beta)(H\beta-h)' + (H\beta-h)(b-\beta)'H'$$

$$+ (H\beta-h)(H\beta-h)' . \qquad (4.3.14)$$

Thus,

$$E[(Hb-h)(Hb-h)'|A_o]P(A_o)$$

$$= H\eta_1 H' + H\eta_2(H\beta-h)' + (H\beta-h)\eta_2' H' + (H\beta-h)(H\beta-h)'[1-g(\theta,\lambda)] ,$$

where η_1 and η_2, as mentioned above, are the expressions found in (4.3.8) and (4.3.9). Substituting the values for η_1 and η_2 into the above expression and simplifying leads to

$$(d) = \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[1-r(\theta,\lambda)]$$

$$+ S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}[1 - s(\theta,\lambda)] \quad . \tag{4.3.15}$$

Summing the four terms (a), (b), (c), and (d) leads to the following:

$$MSE\ (\theta,\lambda) = \sigma^2 S^{-1} + \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[2r(\theta,\lambda)-2 + 1 - r(\theta,\lambda)]$$

$$+ S^{-1}H'[HS^{-1}H]^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}[2s(\theta,\lambda)$$

$$- 2r(\theta,\lambda) + 1 - s(\theta,\lambda)] \quad .$$

That is,

$$MSE\ (\theta,\lambda) = \sigma^2 S^{-1} + \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[r(\theta,\lambda)-1)$$

$$+ S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}$$

$$[1 - 2r(\theta,\lambda) + s(\theta,\lambda)] \quad . \tag{4.3.16}$$

4.4 A Partial Check on the Mean Square Error

When $\lambda = \infty$, $\hat{\beta}$ is chosen with probability 1, $r(\theta,\infty)$ and $s(\theta,\infty)$ both equal zero, so that (4.5.16) becomes

$$MSE(\theta,\infty) = E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$

$$= \sigma^2\{S^{-1}-S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}\} + S^{-1}H'[HS^{-1}H']^{-1}$$

$$(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1} \quad .$$

The risk function of (4.6.1), however, is equivalent to the quadratic risk function, $M(\theta,\lambda)$, defined in (1.6.1), as

$$\begin{aligned} \mathrm{tr}\, X'X[MSE(\theta,\lambda)] &= \mathrm{tr}\, X[E(\beta^*-\beta)(\beta^*-\beta)']X' \\ &= E\,(X\beta^*-X\beta)'(X\beta^*-X\beta) \quad . \end{aligned}$$

Any conclusions drawn from this weighted risk function would be applicable to the predicted value of y and not to the estimated beta vector.

4.6.2 A Sawa-Type Risk Function

It was pointed out in section 2.3 that Sawa and Hiromatsu studied the problem of a linear combination $\underset{\sim}{C}'\underset{\sim}{\beta} = \underset{\sim}{C}_1'\underset{\sim}{\beta}_1 + \underset{\sim}{C}_2'\underset{\sim}{\beta}_2$, of the β's. By a normalizing process involving the supremum of the quadratic risk function over the vector $\underset{\sim}{C}_2$, they arrived at the risk function of (2.3.2), *viz.*

$$R(\theta,\lambda) = h(\theta,\lambda) + \theta[s(\theta,\lambda) - 2h(\theta,\lambda) + 1] \quad . \tag{4.6.2}$$

The supremum they defined would be attained when the $\underset{\sim}{C}_2$ vector had the same direction as the $\underset{\sim}{\beta}_2$ vector, for they made use of the Cauchy-Schwarz inequality

$$|\underset{\sim}{\beta}_2'\underset{\sim}{C}_2| \le ||\underset{\sim}{\beta}_2|| \cdot ||\underset{\sim}{C}_2|| \tag{4.6.3}$$

They then defined a regret function as in (3.2.2), namely

$$REG(\theta,\lambda) = R(\theta,\lambda) - \inf R(\theta,\lambda) \tag{4.6.4}$$

and they sought the optimal value of λ, λ^*, such that

$$REG(\theta,\lambda^*) = \sup_{\theta} REG(\theta,\lambda^*) \quad . \tag{4.6.5}$$

This expression is the same as obtained by Wallace[5] providing a partial check for (4.5.16).

When $\lambda = 0$, b is chosen with probability 1, $r(\theta,0)$ and $s(\theta,0)$ both equal 1, so that (4.5.16) becomes MSE $(\theta,0) = E(b-\beta)(b-\beta)' = \sigma^2 S^{-1}$ as expected.

4.5 Orthogonal Transformations

In section 1.7, a set of orthogonal transformations was defined on the sample, parameter and restriction spaces as follows:

$$Z = XC, \qquad \alpha = C'\beta \qquad \text{and } H_* = HC$$

such that $CC' = I$.

Under this set of transformations, the following can easily be justified:

$$S^{-1} = CC'S^{-1}CC' = C(Z'Z)^{-1}C' \quad , \tag{4.5.1}$$

$$HS^{-1}H' = HCC'S^{-1}CC'H = H_*(Z'Z)^{-1}H_* \quad , \tag{4.5.2}$$

$$\begin{aligned}(H\beta-h)(H\beta-h)' &= (HCC'\beta-h)\;(HCC'\beta-h)' \\ &= (H_*\alpha-h)(H_*\alpha-h)' \quad .\end{aligned} \tag{4.5.3}$$

In section 1.7, it was shown that the bias of y^* and also its quadratic risk function are invariant under the orthogonal transformations defined in that section. Under these same transformations, the bias of β^* would become

[5]Unpublished paper. See footnote 1.

$$\text{bias } \beta^* = CC'S^{-1}CC'H[HCC'S^{-1}CC'H]^{-1}(HCC'\beta-h)(r(\theta,\lambda)-1)$$

$$= C(Z'Z)^{-1}H_*'[H_*(Z'Z)^{-1}H_*']^{-1}(H_*\alpha-h)(r(\theta,\lambda)-1) \ .$$

Clearly, this is not invariant under these transformations.

Let $MSE^*(\theta,\lambda)$ be the transformation of $MSE(\theta,\lambda)$ with the above substitutions made as shown by the statements (4.6.1) through (4.6.3). From the form of $MSE(\theta,\lambda)$ as found in (4.3.16), it follows that

$$C[MSE(\theta,\lambda)]C' = MSE^*(\theta,\lambda)$$

or

$$MSE(\theta,\lambda) = C'[MSE^*(\theta,\lambda)]C \ . \tag{4.5.4}$$

Clearly, $MSE(\theta,\lambda)$ is not invariant under this set of transformations. Whereas a regret function based on the loss function $E(X\beta^*-X\beta)'(X\beta^*-X\beta)$ did not depend on the specific values taken by the S and H matrices, this will not be the case with the loss function $E(\beta^*-\beta)'(\beta^*-\beta)$.

The question remains whether for given values of S and H, a meaningful regret function can be defined in order to obtain an optimal critical value for the prior test. Some possible approaches to this problem are explored in the next section.

4.6 Minimax Regret Functions for the Estimated Beta Vector

4.6.1 A Weighted Risk Function

To overcome the problems mentioned in the previous section, a weighted risk function, $R(\theta,\lambda)$, could be defined by:

$$R(\theta,\lambda) = \text{tr } X'X[MSE(\theta,\lambda)] \tag{4.6.1}$$

and a minimax regret equation set up as in (3.1.3).

This necessitated evaluating $R(\theta,\lambda)$ at the four points

$$R(0,\lambda^*),\ R(0,\infty),\ R(\theta_o,\lambda^*),\ R(\theta_o,0) \quad .$$

But in their formulation $\theta = \frac{\underset{\sim}{\beta}_2'\underset{\sim}{\beta}_2}{\sigma^2}$, so that $\theta=0$ implies that $\underset{\sim}{\beta}_2 = \underset{\sim}{0}$, whereas $\theta=\theta_o$ implies that $\underset{\sim}{\beta}_2 = \underset{\sim}{\beta}_2^o$ where $\underset{\sim}{\beta}_2^o$ is non-zero. It has been shown above that the supremum would only be attained if $\underset{\sim}{C}_2$ had the same direction as $\underset{\sim}{\beta}_2$, and as $\underset{\sim}{C}_2$ is fixed but arbitrary, it is clear that the supremum cannot be attained at both $\theta=0$ and $\theta=\theta_o$. On the other hand, the quadratic risk function of y^*, $M(\theta,\lambda)$ obtained in section 1.6, did not require taking the supremum but gave a value comparable with their $R(\theta,\lambda)$. This indicates that whereas they had tried to find an optimum λ, λ^*, relating to a regret function for the $\underset{\sim}{\beta}$ vector, or at least to linear combinations of the β_i, their arguments actually apply to a regret function involving the predicted value of y.

It may be tempting to set up a Sawa-type risk function in the general context of this paper, although the above comments suggest caution. A risk function, $R(\theta,\lambda)$, could be defined by

$$\begin{aligned} R(\theta,\lambda) &= \sup_{\beta} \operatorname{tr} MSE(\theta,\lambda) \\ &= \sup_{\beta} E\,(\beta^*-\beta)'(\beta^*-\beta) \quad . \end{aligned} \tag{4.6.6}$$

Lancaster [1969, p. 109] gives a theorem showing that for any real kx1 vector γ, and S being a real symmetric matrix, then $\mu_1\gamma'\gamma \le \gamma' S\gamma \le \mu_k\gamma'\gamma$, and, in particular

$$\gamma'\gamma \le \frac{1}{\mu_1}\gamma' S\gamma \quad , \tag{4.6.7}$$

where μ_1, μ_k are, respectively, the smallest and largest eigenvalues of S.

It is clear that the γ vector could be replaced, in turn, by the vectors $(\beta^*-\beta)$, $(\hat{\beta}-\beta)$ or $(b-\beta)$. If expectations are found, then the inequalities remain valid. In the spirit of the risk function of Sawa and Hiromatsu quoted in section 2.3, let

$$R(\theta,\lambda) = \sup_{\beta} E(\beta^*-\beta)'(\beta^*-\beta) = \frac{1}{\mu_1} E(\beta^*-\beta)'S(\beta^*-\beta) \tag{4.6.8}$$

$$= \frac{1}{\mu_1} M(\theta,\lambda) \quad .$$

If the minimax regret condition of (4.6.5) is set up, it is clear that both sides of the equation are identical with the results obtained from the predicted value of y, y^*, but both sides are inflated by the factor $\frac{1}{\mu_1}$. It is clear from the argument above that the $\sup_{\beta}$ tr MSE(θ,λ) cannot, in fact be simultaneously attained for two different values of θ. Any conclusions drawn from this procedure will only strictly be valid for the estimated value of y.

4.6.3 Minimax Regret Condition Based on the Mean Square Error

In the two previous sections, a risk function was defined either as a weighting, or by taking the supremum, of the mean square error of the estimated beta vector. Both of these approaches led to the same critical value given by the predicted value of y.

In this section, a critical value specific to the estimated beta vector will be obtained by taking the trace of the mean square error as the risk function. That is,

$$R(\theta,\lambda) = \text{tr MSE}(\theta,\lambda) = E(\beta^*-\beta)'(\beta^*-\beta) \quad . \tag{4.6.9}$$

In (4.3.16) the mean square error was given by

$$MSE(\theta,\lambda) = \sigma^2 S^{-1} + \sigma^2 S^{-1}H'[HS^{-1}H']^{-1}HS^{-1}[r(\theta,\lambda)-1] \quad (4.6.10)$$
$$+ S^{-1}H'[HS^{-1}H']^{-1}(H\beta-h)(H\beta-h)'[HS^{-1}H']^{-1}HS^{-1}[1-2r(\theta,\lambda) + s(\theta,\lambda)]$$

Let $A = [HS^{-1}H']^{-1}HS^{-2}H'$. The risk function, $R(\theta,\lambda)$, can be written

$$R(\theta,\lambda) = \sigma^2 trS^{-1} + \sigma^2 trA(r(\theta,\lambda)-1) \quad (4.6.11)$$
$$+ (H\beta-h)'A[HS^{-1}H']^{-1}(H\beta-h)[1-2r(\theta,\lambda) + s(\theta,\lambda)] \ .$$

Consider, now, the ratio $\Lambda = \dfrac{(H\beta-h)'A[HS^{-1}H']^{-1}(H\beta-h)}{(H\beta-h)'[HS^{-1}H']^{-1}(H\beta-h)}$. Λ is then of the form $\dfrac{Z' E Z}{Z' F Z}$ and by a simple extension of the theorem quoted in (4.6.7) from Lancaster [1969, p. 109]:

$$\mu_1 \leq \Lambda \leq \mu_m \quad (4.6.12)$$

where μ_i, $i=1,2,\ldots,m$, are the ordered characteristic roots of $EF^{-1} = A$. As $(H\beta-h)'[HS^{-1}H']^{-1}(H\beta-h) = 2\sigma^2 \theta$, the risk function can be simplified further.

$$R(\theta,\lambda) = \sigma^2\{trS^{-1} + trA(r(\theta,\lambda)-1) + 2\theta\Lambda[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \ . \quad (4.6.13)$$

A regret function and minimax regret condition can now be defined in a similar manner as in section 3.1.

Now $R(\theta,0) = \sigma^2 tr\ S^{-1}$, and

$$R(\theta,+\infty) = \sigma^2\{trS^{-1} - trA + 2\theta\Lambda\} \ .$$

The analogs of (3.1.1) and (3.1.2) are, respectively,

$$\inf_{\lambda} R(\theta,\lambda) = \begin{cases} R(\theta,+\infty) & \text{for } \theta < c \\ R(\theta,0) & \text{for } \theta \geq c \end{cases} ,$$

where

$$c = \frac{1}{2}\,\frac{\mathrm{tr}\ A}{\Lambda} \tag{4.6.14}$$

and

$$\mathrm{REG}(\theta,\lambda) = R(\theta,\lambda) - \inf_{\lambda} R(\theta,\lambda) \quad . \tag{4.6.15}$$

For $\theta \geq c$,

$$\mathrm{REG}(\theta,\lambda) = \sigma^2\{\mathrm{tr}A[r(\theta,\lambda)-1] + 2\theta\Lambda[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \quad . \tag{4.6.16}$$

As $r(\theta,\lambda)$ is a probability, and from (3.2.2)

$$s(\theta,\lambda) - r(\theta,\lambda) > 0, \text{ then } [1-2r(\theta,\lambda) + s(\theta,\lambda)] > 0$$

and

$$\sup_{\theta \geq c} \mathrm{REG}(\theta,\lambda) \leq \sup_{\theta \geq \frac{\mathrm{tr}A}{2\mu_m}} \sigma^2\{\mathrm{tr}A[r(\theta,\lambda)-1] + 2\theta\mu_m[1-2r(\theta,\lambda) + s(\theta,\lambda)]\} \tag{4.6.17}$$

It should be noted that the supremum on the right-hand side of (4.6.17) is taken over the larger set of values of θ, $[\frac{\mathrm{tr}A}{2\mu_m}, +\infty)$, which includes the set of values of θ on the left-hand side, $[c, +\infty)$, of (4.6.17).

For $\theta < c$,

$$\mathrm{REG}(\theta,\lambda) = \sigma^2\{\mathrm{tr}\ A\ r(\theta,\lambda) + 2\theta\Lambda[-2r(\theta,\lambda) + s(\theta,\lambda)]\} \quad . \tag{4.6.18}$$

To find the supremum of this quantity, it is necessary to know the sign of the expression, $E = [-2r(\theta,\lambda) + s(\theta,\lambda)]$. As θ tends to infinity, this expression tends to -1. Values of E were computed at 3,960 points for m in the interval [1,120], $n-k$ in [2,120], λ in [0, 3] and θ in $[0, \frac{m}{2}]$.

The results indicated that $E \leq 0$ and is monotonically decreasing with θ to -1, and monotonically increasing with λ to zero.

For the triplet of values $(m, n-k, \lambda)$ considered in Chapter 5, the supremum of $REG(\theta,\lambda)$ for $\theta<c$ will occur when Λ takes the value μ_1, the smallest characteristic root of the matrix A. Again, as c is unknown, the supremum will be taken over the larger set of θ values from 0 to $\frac{trA}{2\mu_1}$. Thus,

$$\sup_{\theta<c} REG(\theta,\lambda) \leq \sup_{\theta<\frac{trA}{2\mu_1}} \sigma^2\{trA.r(\theta,\lambda) + 2\theta\mu_1[-2r(\theta,\lambda) + s(\theta,\lambda)]\}. \quad (4.6.19)$$

The minimax procedure will be to search for an optimal λ, λ_o, which minimizes the expressions in (4.6.17) and (4.6.19). Empirical results will show that as the critical value, λ, increases, then $\sup_{\theta<c} REG(\theta,\lambda)$ decreases but $\sup_{\theta\geq c} REG(\theta,\lambda)$ increases. The optimal minimax regret value λ_o will be that value which equalizes the expressions in (4.6.17) and (4.6.19). Thus, λ_o satisfies the equation

$$\underset{\theta<c}{REG}(\theta_L, \lambda_o) = \underset{\theta\geq c}{REG}(\theta_U, \lambda_o) \quad (4.6.20)$$

where θ_L, θ_U maximize the expressions on the left and right, respectively. An important question arises as to whether it is possible to find such a θ_U which maximizes the quantity on the right of (4.6.17) while at the same time being compatible with the condition that Λ attain its maximum.

Now, Lancaster [1969, p. 110] shows that Λ attains the value μ_m when $H\beta-h = Z_m$, a characteristic vector of the matrix A corresponding to the characteristic root μ_m. This implies that $\theta = \frac{Z_m[HS^{-1}H']^{-1}Z_m}{2\sigma^2}$. No restriction has yet been placed on σ^2 which can take values in the open interval $(0, +\infty)$ so that θ may attain the value of θ_U to maximize the quantity in (4.6.17).

Similarly for (4.6.19), Λ attains the value μ_1 when $H\beta - h = Z_1$ and

$$\theta = \frac{Z_1[HS^{-1}H']^{-1}Z_1}{2\sigma^2} \quad . \tag{4.6.21}$$

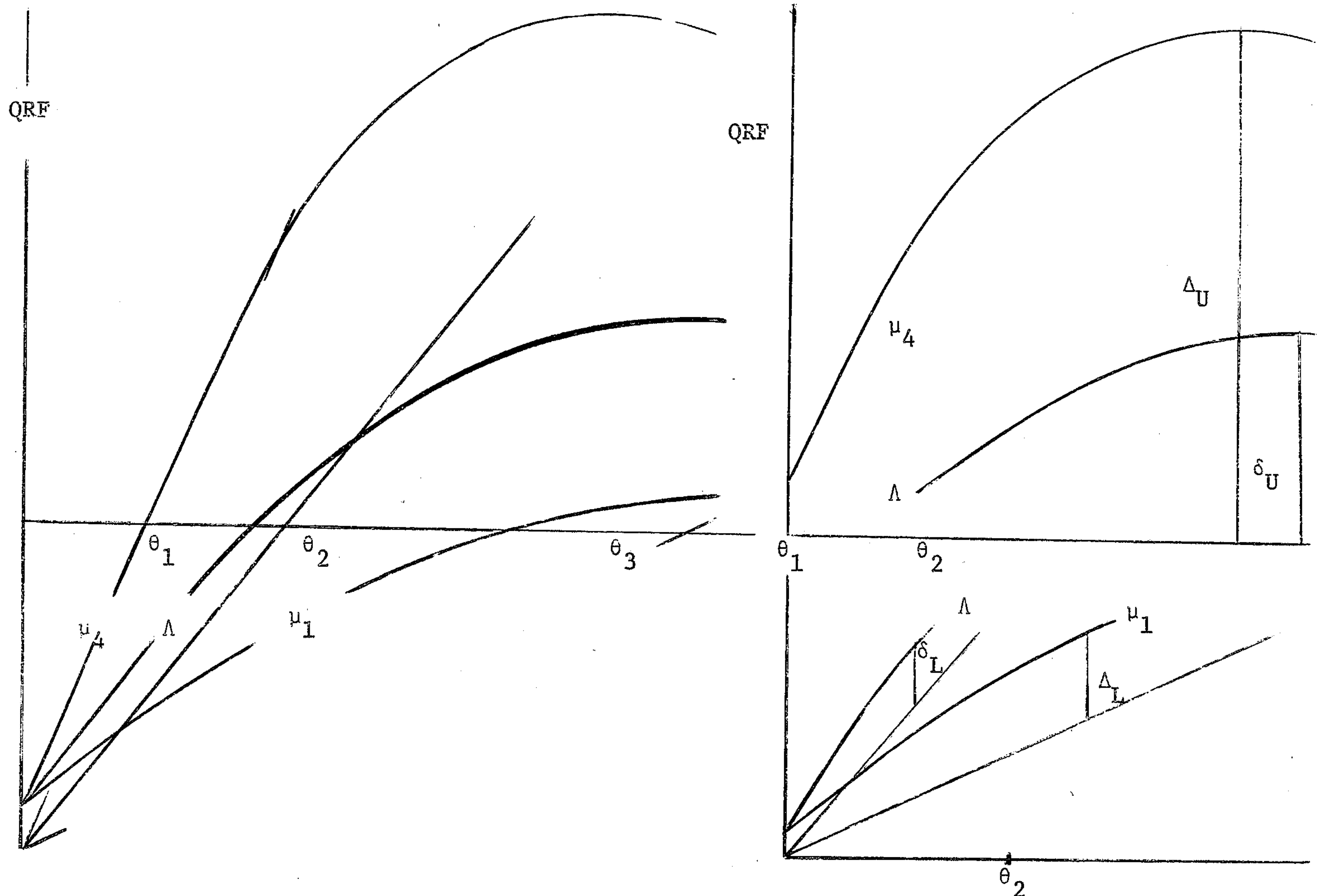
For all values of σ^2, it would be possible to find a θ_L which satisfies both (4.6.19) and (4.6.21).

It is instructive to notice a connection between the regret condition based on the estimated beta vector and the regret condition of Chapter 3 based on the predicted value of y.

Now the trace of a matrix is the sum of its characteristic roots, that is $\operatorname{tr} A = \sum_{i=1}^{m} \mu_i$. If Λ is taken to be the arithmetic mean of the μ_i, that is $\Lambda = \frac{1}{m} \sum_{i=1}^{m} \mu_i$, it is clear that the value of $c = \frac{\operatorname{tr} A}{2\Lambda} = \frac{m}{2}$ and the regret functions of (4.6.16) and (4.6.18) are then equivalent to the regret functions based on the predicted value of y.

This is intuitively appealing as basing a risk function on the predicted value of y is felt to be an averaging process.

The minimax regret condition of this section is illustrated by Figure 4.1. The risk function, $R(\theta, \lambda)$, as found in the expression (4.6.13) contains the unknown variable Λ, which can take on any value between μ_1 and μ_m. For a given value of Λ, the graph of $R(\theta, \lambda)$ lies between the graphs of $R(\theta, \lambda)$ for μ_1 and μ_m, as illustrated by Figure 4.1 (a).



$\theta_1 = \mathrm{tr}A/2\mu_m$

$\theta_2 = \mathrm{tr}A/2\Lambda$

$\theta_3 = \mathrm{tr}A/2\mu_1$

Figure 4.1. The regret condition for the estimated beta vector

In Chapter 3, for $\frac{\Lambda}{trA} = \frac{1}{m}$, the aim was to find a value of the critical point, λ, which equalized the distances δ_L and δ_U. For unknown Λ these distances cannot be evaluated, but in Figure 4.1 (b) and (c), it can be seen that these distances are maximized by the distances Δ_L and Δ_U. The minimax regret criterion minimizes these maximum deviations in mean square error, and this occurs when $\Delta_L = \Delta_U$.

5. RESULTS AND CONCLUSIONS

5.1 Computer Procedure Followed to Obtain Critical Values Based on the Predicted Value of y

It was shown in section 3.1 that the optimal critical value, λ^*, of the prior test of estimation satisfies equation (3.1.4), viz.

$$\underset{\theta<\frac{m}{2}}{\mathrm{REG}}(\theta_L,\lambda^*) = \underset{\theta\geq\frac{m}{2}}{\mathrm{REG}}(\theta_U,\lambda^*) \tag{5.1.1}$$

where θ_L, θ_U are the values of θ which maximize the quantities on the left and right, respectively.

The procedure followed was to give values to m, the numerator degrees of freedom, n-k, the denominator degrees of freedom, and λ, the critical value of the prior F test.

To find θ_U, a computer search is carried out by increasing θ in small increments from a starting value of $\frac{m}{2}$. The slope of the quadratic risk function of y^*, $M(\theta,\lambda)$, is calculated at each θ value. In this way, a small interval, I, along the θ axis is located such that the slope of $M(\theta,\lambda)$ changes from positive to negative in the interval I. By successively reducing the size of I, the optimal value for θ, θ_U, is obtained. In practice, this value is assumed to be reached when successive iterations differ by less than 10^{-4}.

In a similar way, the optimal value, θ_L, is obtained for $\theta < \frac{m}{2}$.

Denote the left- and right-hand side of (5.1.1) by δ_L and δ_U, respectively. If these differ by less than 10^{-7}, then the current value of λ is taken to be the optimal value, λ^*. Otherwise, λ is increased or decreased according to whether δ_L is less than or more than δ_U, respectively. That this procedure leads to an optimal value of λ confirms that for increasing λ, $\underset{\theta<\frac{m}{2}}{\mathrm{REG}(\theta_L,\lambda)}$ decreases but $\underset{\theta\geq\frac{m}{2}}{\mathrm{REG}(\theta_U,\lambda)}$ increases.

As a check that the maximum regret for $\theta \geq \frac{m}{2}$ occurs at θ_U, the slope of the regret function is checked for each λ at twenty points between θ_U and 20m. For the optimal critical point, λ^*, θ_U is found to be less than 2m so that checking that the slope is negative for values of θ up to 20m provides strong confirmation that the maximum regret occurs at θ_U.

Sawa and Hiromatsu [1971] gave values of λ^* for m=1. Their values agreed closely with those of Table 5.1, except that their values were generally lower by 0.001. These authors, however, assumed that the supremum of the regret function for $\theta < \frac{m}{2}$ occurred at $\theta = 0$. While this is true for the case of m=1, this does not hold for larger values of m. Their assumption actually leads to smaller values of λ^* particularly for large denominator degrees of freedom.

Two subroutines were used for the non-central F. Both of these were written by Mr. James Goodnight for generating the tables in Goodnight and Wallace.[6] One of these subroutines uses a quick approximation by means of a central F, while the other uses an iterative procedure until a given required accuracy is obtained. The approximate subroutine was used to generate optimal values for λ for the numerator and

---

[6] James Goodnight and T. D. Wallace, N. C. State University, Raleigh, N. C., Operational techniques and tables for making weak MSE tests for restrictions in regressions, to be published in Econometrica.

denominator degrees of freedom listed in Table 5.1. The results were checked with the more accurate subroutine for m = 1,2,4,8, and the discrepancies between the values of λ^* obtained by the two subroutines were less than 5.5 percent, 2 percent, 0.55 percent, 0.13 percent for m = 1,2, 4, and 8, respectively. As the discrepancies are small for higher values of m, the values listed in Table 5.1 were obtained from the accurate subroutine for m = 1,2,4,8 and from the fast approximation for the remainder of the table.

Certain monotone properties of θ_L, θ_U and λ^* emerged which were used to limit the area of search when the slower, more accurate subroutine was used to generate the final tables of results for m = 1,2,4,8. These properties are detailed in the next section.

5.2 Results Based on the Predicted Value of y

From Table 5.1, it is clear that the outstanding property of the optimal critical value, λ^*, is that it does not vary much from the value 2. When m, the numerator degrees of freedom, is 1, λ^* only decreases by about 0.1 as the denominator degrees of freedom increases from 2 to 120. For large m, this decrease is more pronounced and when m = 120, λ^* decreases by about 0.3 as the denominator degrees of freedom increases from 2 to 120. For each value of m, λ^* decreases with the denominator degrees of freedom, which helps considerably in the search technique for when the numerator degrees of freedom are increased, the previous value of λ^* is used as an upper limit for the starting interval. For a given value of the denominator degrees of freedom, n-k, λ^* increases with m, the numerator degrees of freedom.

Table 5.1. Optimal values for the predicted value of y

m	n-k	λ^*	$\alpha(F)$	θ_L	θ_U
1	2	1.972	35.4	0.0	2.0
	4	1.922	26.8	0.0	1.9
	8	1.898	20.8	0.0	1.9
	16	1.887	17.3	0.0	1.8
	24	1.883	15.9	0.0	1.8
	60	1.878	14.3	0.0	1.8
	120	1.876	13.7	0.0	1.7
2	2	2.097	34.8	0.0	3.9
	4	2.006	25.8	0.0	3.2
	8	1.952	19.5	0.0	2.8
	16	1.922	15.6	0.0	2.7
	24	1.911	14.1	0.1	2.6
	60	1.897	12.3	0.1	2.5
	120	1.892	11.6	0.1	2.5
4	2	2.185	34.7	0.0	7.1
	4	2.058	25.3	0.0	5.5
	8	1.984	19.0	0.4	4.7
	16	1.953	15.1	0.4	4.3
	24	1.944	13.6	0.5	4.2
	60	1.931	11.7	0.5	4.0
	120	1.931	11.0	0.5	3.9
8	2	2.237	34.6	0.0	13.6
	4	2.093	24.9	0.3	10.3
	8	2.019	17.1	1.1	8.5
	16	1.988	11.6	1.6	7.6
	24	1.979	9.4	1.7	7.2
	60	1.972	6.6	2.0	6.8
	120	1.969	5.7	2.1	6.6
16*	2	2.265	34.5	0.0	26.3
	4	2.119	24.3	1.0	19.8
	8	2.043	15.4	2.5	16.1
	16	2.008	8.7	3.6	14.0
	24	1.997	6.1	4.1	13.2
	60	1.986	3.0	4.8	12.1
	120	1.983	2.0	5.0	11.7
24*	2	2.275	34.6	0.0	39.3
	4	2.128	24.1	1.6	29.4
	8	2.052	14.7	4.0	23.7
	16	2.016	7.6	5.7	20.4
	24	2.005	4.8	6.5	19.1
	60	1.993	1.7	7.7	17.4
	120	1.989	0.9	8.1	16.7

Table 5.1 (continued)

m	n-k	λ^*	$\alpha(F)$	θ_L	θ_U
60*	2	2.286	34.7	0.0	97.8
	4	2.140	23.9	4.4	72.5
	8	2.064	13.7	10.5	58.0
	16	2.027	6.0	15.2	49.3
	24	2.015	3.1	17.4	45.8
	60	2.002	0.5	20.9	40.8
	120	1.997	0.1	22.5	38.7
120*	2	2.290	34.8	0.0	195.2
	4	2.144	23.8	9.0	144.5
	8	2.069	13.4	21.4	115.2
	16	2.032	5.4	31.1	97.4
	24	2.020	2.5	35.6	90.2
	60	2.005	0.2	43.1	79.6
	120	2.001	0.1	46.8	74.9

m = number of restrictions

n = number of points in the sample space

k = number of parameters

λ^* = critical value of the prior test of estimation

θ_L, θ_U = optimal lower, and upper, theta values

$\alpha(F)$ = percentage alpha level for the central F

*These values were computed using the approximate subroutine.

Although the optimal values of the critical points, λ^*, of the prior test of estimation fluctuate only slightly over the whole range of numerator and denominator degrees of freedom, the corresponding alpha levels vary considerably as shown by Table 5.1. For denominator degrees of freedom of at least four, the percentage alpha level decreases monotonically with numerator and denominator degrees of freedom. This reflects the fact that the significance points of a central F decrease monotonically with numerator degrees of freedom, provided that the denominator degrees of freedom are at least four.

If a prior test of estimation is performed with alpha at the 5 percent level, then the unrestricted least squares estimate will be chosen with probability of 5 percent. Clearly, from Table 5.1, the minimax regret function, as compared to a 5 percent F test, generally increases the probability that the unrestricted least squares estimate will be chosen, and in many cases this probability is greater than 15 percent. For large m and n-k, however, the alpha level decreases until it is only 0.1 percent for m and n-k both equal to 120.

Table 5.1 also gives optimum values, θ_L, θ_U, of the non-centrality parameter. As the numerator degrees of freedom, m, increase, both θ_L and θ_U increase. As the denominator degrees of freedom, n-k, increase, θ_L increases but θ_U decreases.

For small m and n-k, θ_L is zero. This occurs as $r(\theta,\lambda)$ and $s(\theta,\lambda)$ are relatively large for small values of m and n-k. This would be expected from the definition of $r(\theta,\lambda)$ as found in (3.2.4) and from the fact that for a given alpha level, the critical points of a non-central F distribution are monotonically decreasing with numerator and denominator degrees of freedom as can be seen by the tables in

Goodnight and Wallace[7]. When $r(\theta,\lambda)$ and $s(\theta,\lambda)$ are large, it is found that $[-2r(\theta,\lambda) + s(\theta,\lambda)]$ is large and negative so that the maximum of the regret function from the left-hand side of (3.1.6) occurs when θ_L is zero.

For large degrees of freedom in the numerator and denominator, θ_U approaches, from above, half the numerator degrees of freedom. This can be seen in Figure 5.1 as the distance δ_U moves to the point $\theta=m/2$ as m increases from 2 to 120. It should be noted that these graphs are scaled so that both the quadratic risk function and the values of theta are expressed in terms of m. It can be seen also that as m increases the shape of the graph changes so that the maximum regret, δ_L, for $\theta < m/2$ occurs at $\theta = 0$ for $m = 2$ but moves away from zero for $m = 8$ and $m = 120$.

5.3 Computer Procedure for the Case of the Estimated Beta Vector

The procedure followed to obtain optimum values based on the beta vector was similar to that used for the predicted value of y. The only difference was that, instead of equalizing the distances δ_L and δ_U of Figure 3.1, the aim was to equalize the distances Δ_L and Δ_U of Figure 4.1.

As Δ_L depends on the smallest root of A, μ_1, and Δ_U depends on the largest root, μ_m, then extensive tables would be needed to cover the possible values that μ_1 and μ_m could take. For a given set of data, S and H would be fixed and known so that the matrix $A = [HS^{-1}H']^{-1}HS^{-2}H'$ could be evaluated and the roots μ_1 and μ_m obtained. Denote $\frac{\mu_i}{trA}$ by μ_i^*. For the case of the predicted value of y of each of the μ_i^* equal $\frac{1}{m}$. For

[7] See footnote 6.

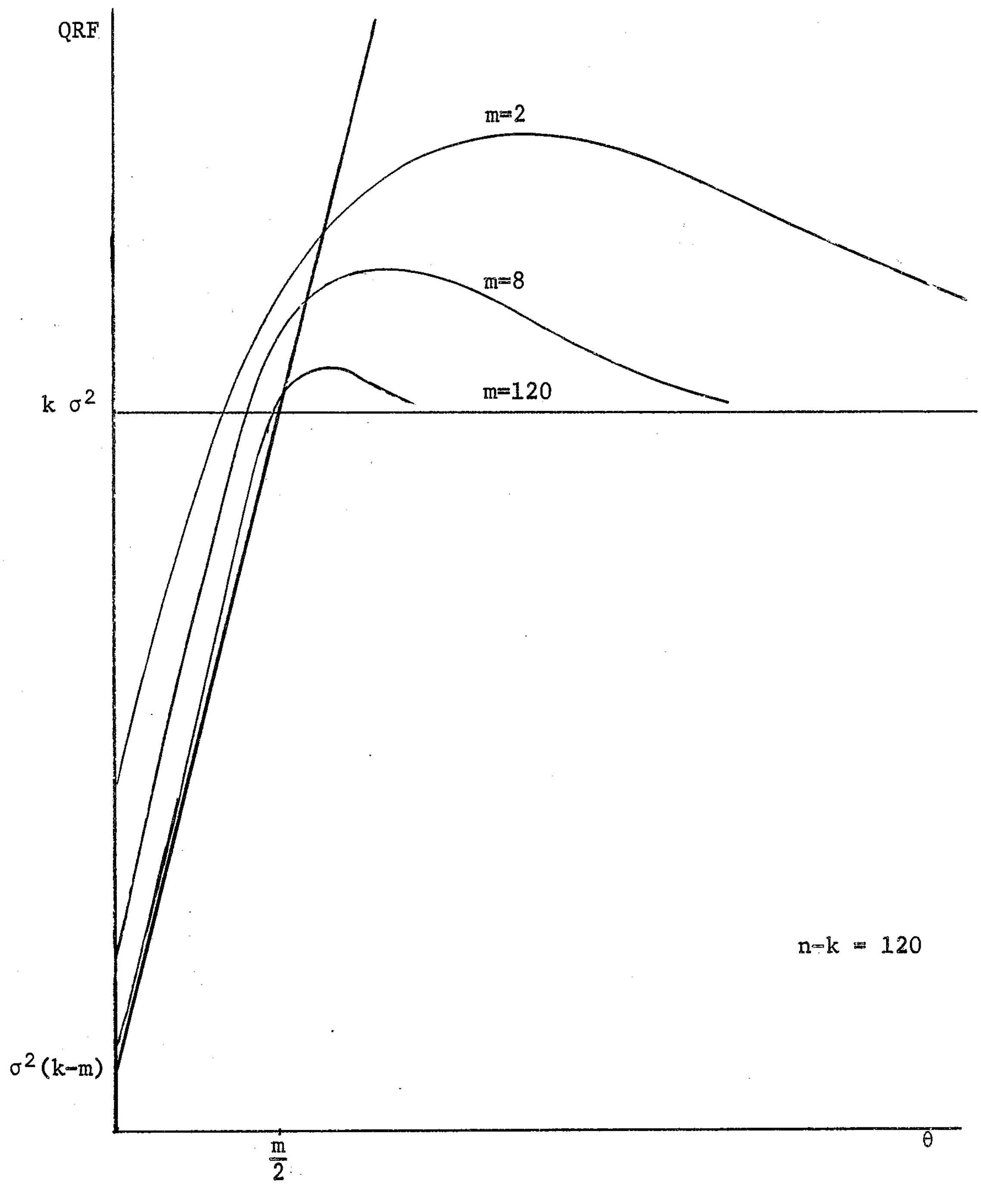


Figure 5.1. The quadratic risk function of y for optimal critical values

the estimated beta vector, $0<\mu_1^* \leq \frac{1}{m}$ and $\frac{1}{m} \leq \mu_m^* < 1$. As the μ_i^* sum to unity, then for a given value of μ_1^*, μ_m^* satisfies the more restrictive inequality

$$\frac{1-\mu_1^*}{(m-1)} \leq \mu_m^* \leq 1-(m-1)\mu_1^* \quad . \tag{5.3.1}$$

In the next section, optimal values will be found for m=4 and n-k = 8 and 60, and for selected values of μ_1^* and μ_4^* satisfying the conditions

$$0 < \mu_1^* \leq 0.25 \quad \text{and} \tag{5.3.2}$$

$$\frac{1-\mu_1^*}{3} \leq \mu_4^* \leq 1-3\mu_1^* \quad . \tag{5.3.3}$$

5.4 Results Based on the Estimated Beta Vector

From Table 5.2, it can be seen that as μ_1, the smallest characteristic root of A decreases, or, in other words, as μ_1^* decreases the optimal critical value, λ_o, increases. This would be expected from (4.6.19) for $[1-2r(\theta,\lambda) + s(\theta,\lambda)]$ being negative implies that the regret for $\theta < c$ increases as μ_1 decreases. This means that a larger value of λ must be chosen to reduce this regret so that it equals the regret for $\theta \geq c$. To illustrate that λ increases as μ_1^* decreases, consider the triplet of values of $(\mu_1^*, \mu_4^*, \lambda_o) = (0.20, 0.30, 1.841)$ and $(0.10, 0.30, 1.937)$.

Consider the regret for $\theta \geq c$, as found in (4.6.17). This involves the expression $2\theta\, \mu_m[1-2r(\theta,\lambda) + s(\theta,\lambda)]$, and as the term in brackets is non-negative, an increase in μ_m implies an increase in the regret, and a smaller value of λ must be chosen to equalize this regret with the regret for $\theta < c$. Thus, an increase in μ_m implies an increase in λ_o,

Table 5.2. Optimal values for the estimated beta vector*

	μ_1^*	μ_4^*	λ_o	$\alpha(F)$	θ_L	θ_U
m=4, n-k=8	0.25	0.25	1.997	18.7	0.4	4.6
	0.20	0.30	1.841	21.4	0.5	4.2
		0.35	1.706	24.0	0.4	3.9
		0.40	1.599	26.4	0.3	3.6
	0.15	0.35	1.730	23.5	0.7	3.9
		0.45	1.526	28.2	0.5	3.4
		0.55	1.384	32.1	0.4	3.1
	0.10	0.30	1.937	19.7	1.4	4.3
		0.40	1.663	25.0	1.2	3.7
		0.50	1.481	29.4	1.0	3.3
		0.60	1.351	33.1	0.8	3.0
		0.70	1.251	36.3	0.7	2.8
	0.05	0.35	1.930	19.8	2.8	4.2
		0.45	1.672	24.8	2.4	3.6
		0.55	1.495	29.0	2.2	3.2
		0.65	1.365	32.7	1.9	2.9
		0.75	1.264	35.9	1.8	2.7
		0.85	1.184	38.7	1.6	2.6
m=4, n-k=60	0.25	0.25	1.951	11.2	0.7	3.9
	0.20	0.30	1.822	13.5	0.8	3.5
		0.35	1.698	16.1	0.7	3.2
		0.40	1.599	18.5	0.6	3.0

Table 5.2 (continued)

	μ_1^*	μ_4^*	λ_o	$\alpha(F)$	θ_L	θ_U
m=4, n-k=60						
	0.15	0.35	1.734	15.3	1.0	3.3
		0.45	1.543	20.0	0.8	2.9
		0.55	1.410	24.0	0.7	2.6
	0.10	0.30	1.947	11.3	1.8	3.7
		0.40	1.688	16.3	1.5	3.1
		0.50	1.516	20.7	1.3	2.8
		0.60	1.391	24.7	1.1	2.5
		0.70	1.296	18.1	1.0	2.4
	0.05	0.35	1.968	10.9	3.1	3.6
		0.45	1.718	15.6	2.7	3.1
		0.55	1.546	19.9	2.4	2.8
		0.65	1.419	23.7	2.2	2.5
		0.75	1.321	27.1	2.0	2.4
		0.85	1.242	30.2	1.8	2.2

m = number of restrictions

n = number of points in the sample space

k = number of parameters

λ_o = critical value of the prior test of estimation

θ_L, θ_U = optimal lower, and upper, theta values

$\alpha(F)$ = percentage alpha level for the central F

$\mu_i^* = \frac{\mu_i}{trA}$, i = 1,4, where μ_i are the ordered roots of the matrix A

*These values were computed from the approximate subroutine.

and an example of this can be seen in the table when $\mu_1^* = 0.10$, as μ_4^* increases from 0.30 to 0.70 then λ_o decreases from 1.937 to 1.251.

It seems clear from the table that the value of μ_4^* has more influence than μ_1^* on the value of λ_o. Consider m=4, n-k = 60, and μ_4^* = 0.35. As μ_1^* increases threefold from 0.05 to 0.15, λ_o only decreases from 1.968 to 1.734. When $\mu_1^* = 0.10$, however, and μ_4^* is doubled from 0.30 to 0.60, λ_o decreases considerably from 1.947 to 1.391.

5.5 Percentage Alpha Levels if λ^* Is a Critical Point of the Non-Central F

Wallace[8] proposed other criteria for linear restrictions in regression based on whether the mean square error of the unrestricted estimator was less than the mean square error of the unrestricted estimator. His "Strong MSE" criterion assumed that the test statistic, u, is distributed as a non-central F with θ, the non-centrality parameter, equal to 1/2. On the other hand, his "Weak MSE" criterion assumes that $u \sim F'(m/2)$. One motivation for using these criteria is that they increase the probability of choosing the restricted estimators over the ordinary F test as the restricted estimators have smaller variance even though they are biased under the alternate hypothesis. In section 2.5, it was noted that the minimax regret function in general chooses the unrestricted estimators with higher probability than the usual 5 percent level F test. The percentage alpha levels in the Tables 5.3 and 5.4 show the percentage probability of choosing the unrestricted estimator assuming u is distributed as a non-central F with 1/2 or m/2 as non-centrality parameter. These tables re-emphasize that the Wallace

Table 5.3. Alpha percent level if λ^* is from the non-central F with non-centrality = 1/2[a]

n-k \ m	1	2	4	8	16	24	60	120
2	45.5	42.4	39.7	37.8	36.2	35.7	35.2	35.0
4	41.5	36.7	32.2	28.8	26.3	25.5	24.4	24.0
8	39.2	33.2	26.9	21.3	17.5	16.1	14.3	13.7
16	38.0	31.2	23.3	15.8	10.7	8.8	6.4	5.6
24	37.6	30.5	21.8	13.6	7.9	5.8	3.4	2.6
60	37.1	29.6	20.0	10.4	4.2	2.3	0.5	0.2
120	37.0	29.3	19.3	9.3	3.0	1.3	0.1	0.1

[a]Table gives $100\ \alpha = 100 \int_{\lambda}^{\infty} g(F'(1/2))\ dF'(1/2)$.

Table 5.4. Alpha percent level if λ^* is from the non-central F with non-centrality = m/2[a]

n-k \ m	1	2	4	8	16	24	60	120
2	45.5	51.0	54.4	56.3	57.0	57.3	59.7	57.8
4	41.5	46.9	50.7	53.1	54.1	54.6	55.1	55.3
8	29.2	44.5	48.5	50.9	52.2	52.7	53.4	53.6
16	38.0	43.2	47.0	49.2	50.6	51.2	52.1	52.4
24	37.6	42.7	46.3	48.5	49.8	50.5	51.5	51.9
60	37.1	42.1	45.6	47.3	48.6	49.3	50.4	50.9
120	37.0	41.9	45.3	46.9	48.0	48.7	49.8	50.3

[a]Table gives $100\ \alpha = 100 \int_{\lambda}^{\infty} g(F'(m/2))\ dF'(m/2)$.

criteria increase the probability of choosing restricted estimators except when the non-centrality is 1/2 and m and n are large.

5.6 Efficiency of the Estimators Based on the Optimal Critical Values

The quadratic risk function for the predicted value of the dependent variable has been considered in some detail in Chapter 3, and in section 5.2, graphs of the quadratic risk function were displayed for three values of m, the number of restrictions on the parameter space. Another way of considering the properties of the quadratic risk function is afforded by the concept of relative efficiency.

In Figure 5.2, graphs are drawn for three values of m for the efficiency of the estimator $X\beta^*$ relative to the least squares estimator, Xb. This relative efficiency will be designated by R.E. $(\lambda^*, 0)$ and defined by

$$\text{R.E.}\ (\lambda^*, 0) = \frac{\text{q.r.f. for } \lambda=\lambda^*}{\text{q.r.f. for } \lambda=0}, \text{ where } \lambda^* \text{ is the optimal}$$

value of λ obtained by the minimax regret condition. When the non-centrality parameter, θ, is zero then R.E. $(\lambda^*, 0)$ takes on the value $\frac{k}{k-m+mr(0,\lambda)}$, where $r(0,\lambda) = \Pr\{F(m+2,n-k) \geq \frac{m}{m+2}\lambda\}$. From standard F tables, it is found that this probability decreases as m increases. This is reflected in the graphs in that, for $\theta = 0$, R.E. $(\lambda^*, 0)$ is large when m = 120, and decreases as m decreases.

The graphs are scaled as the non-centrality parameter, θ, is expressed as a multiple of the number of restrictions, m. Even with this scaling factor, the relative efficiency with m large is everywhere greater than the relative efficiency for small n. In fact,

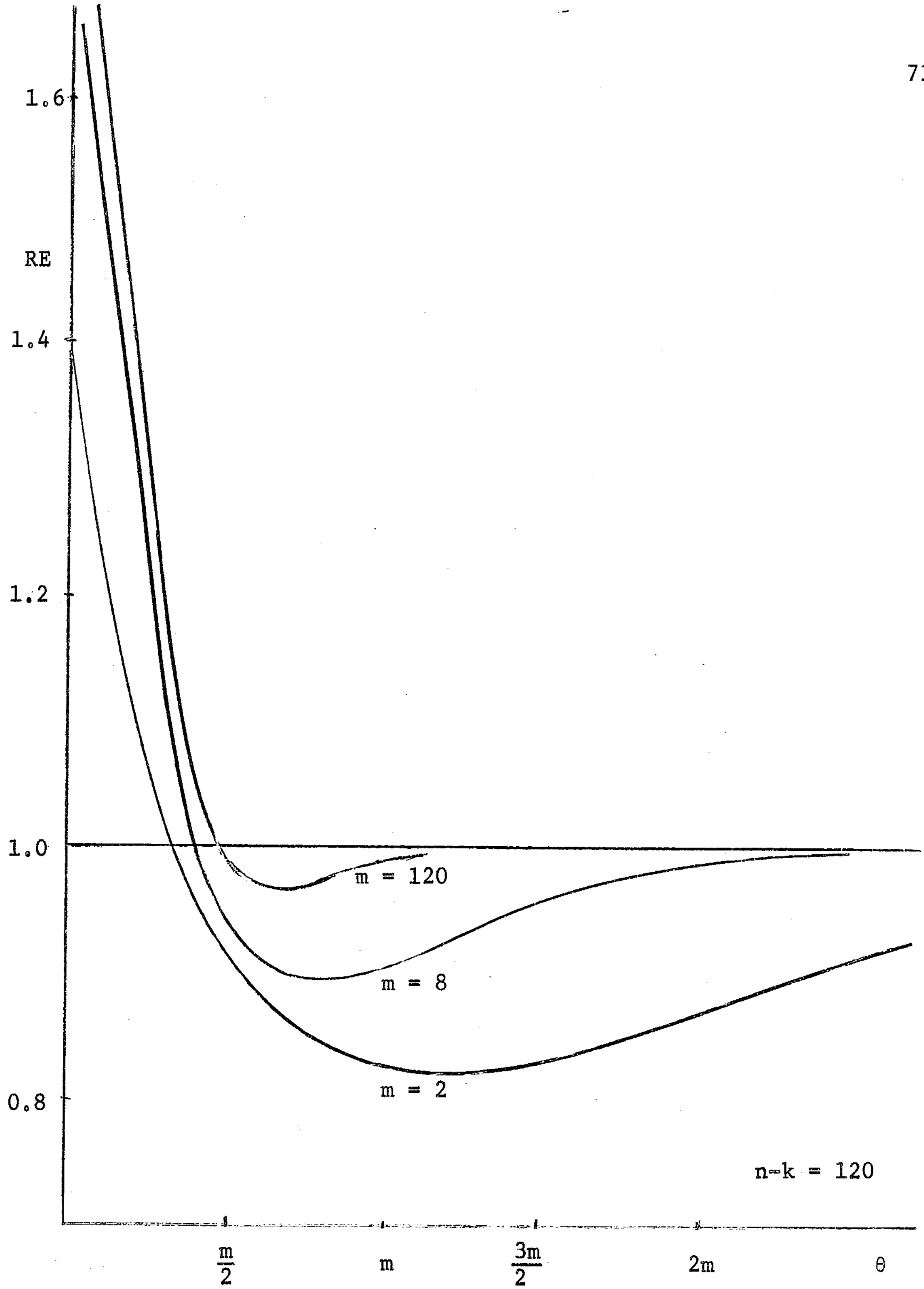


Figure 5.2. Efficiency of the estimator $X\beta^*$ relative to the least squares estimator, Xb

for m = 120, the relative efficiency only falls below 1 for a relatively small interval of θ. Of course, the shape of the curve of R.E. $(\lambda^*, 0)$ will alter as the number of parameters, k, is altered. This can be compared with the graphs of the quadratic risk function as in Figure 5.1 where altering the value of k will only effect a translation of the curve but its shape will be unaffected. As k increases both numerator and denominator of R.E. $(\lambda^*, 0)$ increase so that the value of R.E. $(\lambda^*, 0)$ becomes closer to unity for all values of θ, or, in other words, the curves tend to collapse towards the line R.E. $(\lambda^*, 0) = 1$.

In Figure 5.3, the efficiency of the estimator using the optimal critical value, λ^*, relative to that of the 5 percent F value is plotted for the same three values of m. This could be designated as

$$\text{R.E. } (\lambda^*, \lambda_{0.05}) = \frac{\text{q.r.f. for } \lambda=\lambda^*}{\text{q.r.f. for } \lambda=\lambda_{0.05}} .$$

The curves reflect the fact that, except for large values of m and n, the critical values are smaller than the 5 percent F values. Consequently, the q.r.f. using λ^* is larger than the q.r.f. using $\lambda_{0.05}$ when θ is small, but for larger θ, and certainly when θ is greater than m/2, the opposite is true. Again, as k, the number of parameters, increases, R.E. $(\lambda^*, \lambda_{0.05})$ becomes closer to unity.

5.7 Maximizing the Minimum Relative Efficiency

In the previous section, the efficiency of the estimator, $X\beta^*$, was compared with the ordinary least squares estimator and the traditional two-stage estimator at the 5 percent level. A more interesting measure is the efficiency relative to the "best" estimator which is the restricted estimator, $X\hat{\beta}$, when θ is in the interval $[0, \frac{m}{2}]$, and

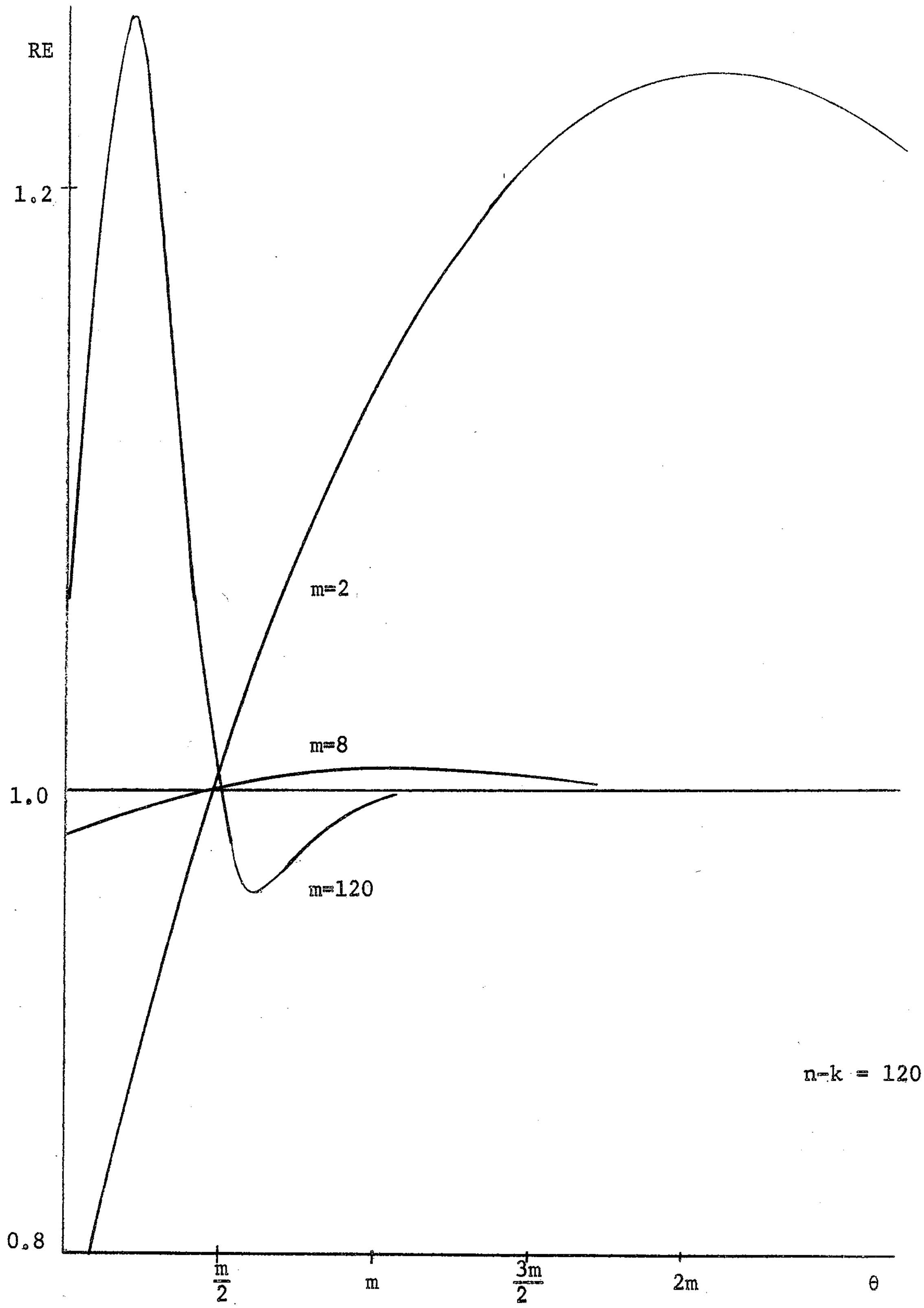


Figure 5.3. Efficiency of the estimator $X\beta^*$ relative to the estimator based on the 5 percent F value

the least squares estimator, Xb, when θ is in $(\frac{m}{2}, +\infty)$. As the quadratic risk function for $X\beta^*$, Xb and $X\hat{\beta}$ are given respectively by:

$$M(\theta,\lambda) = \sigma^2\{k-m+mr(\theta,\lambda) + 2\theta[1-2r(\theta,\lambda) + s(\theta,\lambda)]\}$$

$$M(\theta,0) = \sigma^2 k$$

$$M(\theta, +\infty) = \sigma^2\{k-m+2\theta\}$$

The relative efficiency is then defined by

$$R.E.(\theta,\lambda) = \text{Relative Efficiency} = \begin{cases} \dfrac{M(\theta,+\infty)}{M(\theta,\lambda)} & \text{when } \theta \text{ is in } [0, \frac{m}{2}] \\ \dfrac{M(\theta,0)}{M(\theta,\lambda)} & \text{when } \theta \text{ is in } (\frac{m}{2}, +\infty) \end{cases} \tag{5.7.1}$$

In a similar approach to that of the minimax regret function of Chapter 3, a maximin regret function based on this relative efficiency could be used to determine an optional value of λ, the critical value of the prior test of estimation. This maximin approach would be to find the critical value, λ_o, which satisfies the condition

$$\inf_{\theta<\frac{m}{2}} R.\ E.\ (\theta,\lambda_o) = \inf_{\theta\geq\frac{m}{2}} R.\ E.\ (\theta,\lambda_o) \quad , \tag{5.7.2}$$

where λ_o is in the interval $[0, +\infty]$.

It can be seen from Table 5.5 and Figure 5.4 that this criterion leads to a slightly higher critical value and lower alpha level than the minimax criterion. As the total number of parameters, k, increases, however, the critical value decreases slightly.

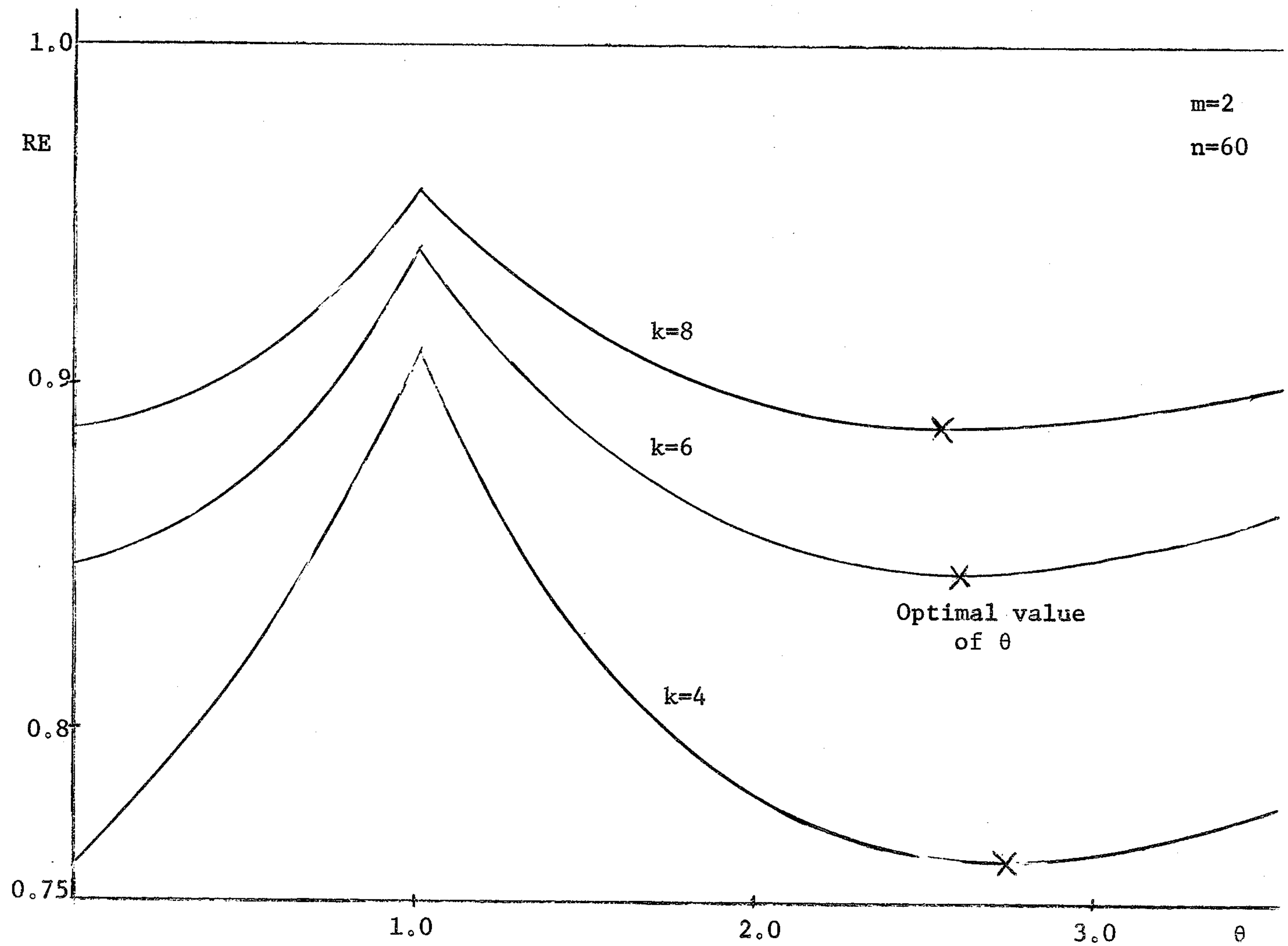


Figure 5.4. Efficiency of $X\beta^*$ relative to the "best" estimator

Table 5.5. Critical values obtained by maximizing the minimum relative efficiency of the estimator, $X\beta^*$, when the number of restrictions is two

n-k	k	λ^*	θ_U	$\alpha(F)$	$\alpha(F'(\frac{1}{2}))$	$\alpha(F'(\frac{m}{2}))$
2	4	3.079	4.7	24.5	33.2	40.9
	6	2.647	4.3	27.4	36.7	44.9
	8	2.487	4.1	28.7	38.2	46.6
4	4	2.782	3.7	17.4	27.2	36.2
	6	2.452	3.4	20.1	30.7	40.2
	8	2.327	3.3	21.3	32.2	42.0
8	4	2.606	3.1	13.3	23.6	33.5
	6	2.336	2.9	15.7	27.0	37.5
	8	2.231	2.9	16.8	28.5	39.2
16	4	2.508	2.9	11.1	21.6	32.0
	6	2.270	2.7	13.4	24.9	36.0
	8	2.176	2.6	14.4	26.4	37.7
24	4	2.473	2.8	10.3	21.0	31.6
	6	2.246	2.6	12.5	24.2	35.5
	8	2.157	2.6	13.5	25.6	37.2
60	4	2.429	2.7	9.4	20.2	31.0
	6	2.217	2.5	11.5	23.3	34.9
	8	2.132	2.5	12.5	24.7	36.5
120	4	2.415	2.6	9.1	19.9	30.8
	6	2.207	2.5	11.2	23.0	34.7
	8	2.124	2.4	12.1	24.4	36.3

Table 5.5 (continued)

---

n = number of points in the sample space

k = number of parameters

λ^* = critical value for the prior test of estimation

θ_U = optimal theta value

$\alpha(F)$ = percentage alpha level for the central F

$\alpha(F'(\frac{1}{2}))$ = percentage alpha level for the non-central F with non-centrality = $\frac{1}{2}$

$\alpha(F'(\frac{m}{2}))$ = percentage alpha level for the non-central F with non-centrality = $\frac{m}{2} = 1$

The lower optimal value of θ, θ_L, was found to be zero for each of the three cases considered. The upper optimal value of θ, θ_U, is the same value, for a given λ, which maximizes the regret function of Chapter 3.

5.8 Conclusions

In the context of a general linear model with linear restrictions, the traditional prior test of estimation leads to a choice between the restricted or unrestricted estimators. The resulting composite estimator is biased.

This study has led to explicit expressions for the bias function of the composite estimator, $X\beta^*$, in the case of estimating the predicted value of the dependent variable, y. As expected, the value of this bias function is found to be always less in absolute value than that of the restricted estimator, while the unrestricted estimator is, of course, unbiased.

Explicit expressions were also obtained for the quadratic risk function for the predicted value of y. This risk function was shown to reach a maximum value when the non-centrality parameter, θ, takes the value θ_U, dependent on the numerator and denominator degrees of freedom and the critical value, λ, of the prior test. Except for some interval of θ between one-fourth of the numerator degrees of freedom and θ_U, the graph of the quadratic risk function lies between that of the restricted and unrestricted estimators.

A minimax regret function based on the quadratic risk function of $X\beta^*$ was then proposed to determine the optimum value of the critical value, $\lambda=\lambda^*$. This value, λ^*, was found to be invariant under orthogonal transformations of the sample, parameter and restriction spaces. This is to be expected as the bias and quadratic risk function of $X\beta^*$ are invariant under these transformations.

From the empirical results, it is clear that the minimax regret function is generally more conservative in choosing the biased restricted estimator than the traditional F test at (say) the 5 percent level. This may be expected as no prior information is assumed on the distribution of the non-centrality parameter and large values of this non-centrality parameter imply large values of the bias and, hence, of the quadratic risk of the restricted estimator.

Instead of the minimax criterion to set the value of λ, a similar criterion could be used based on the efficiency of the estimator relative to the "best" estimator, which is the ordinary least squares estimate when the non-centrality parameter θ is in the interval $(\frac{m}{2},+\infty)$, but is the restricted estimator for θ in the interval $[0, \frac{m}{2}]$. The value of λ, λ_o, is sought which maximizes the minimum efficiency of the estimator. In general, this procedure leads to somewhat larger values of λ, and smaller alpha levels, than does the minimax procedure. The optimum λ values, λ_o, do depend on k, the total number of parameters, in this case and decrease as k increases.

Attention is then focused on the estimate, β^*, of the beta vector, and expressions are found for the bias and the mean square error of β^*. It is found that these expressions are not invariant to orthogonal transformations, so that it would not be possible to find an optimum value of λ which did not depend on the design matrix $S = X'X$, or the matrix H which determines the restrictions on the beta vector. In a particular case when S and H are known, then an optimum value of λ, λ_o, can be found, and it depends on the smallest, μ_1, and the largest, μ_m, characteristic roots of the matrix $A = [HS^{-1}H']^{-1}HS^{-2}H'$.

The empirical results show that λ_o decreases as μ_m or as μ_1 increases. On the other hand, it is shown that the minimax condition for the predicted value of y depends on an averaging process and involves a constant factor, the reciprocal of the number of restrictions.

The main value of the minimax regret condition is that it does not depend on an arbitrary level of the prior F test. In the absence of other information about the non-centrality parameter, and hence of the proposed restrictions, it leads to an estimator with a small quadratic risk function over the whole range of the non-centrality parameter.

6. LIST OF REFERENCES

Ashar, V. G. 1970. On the use of preliminary tests in regression analysis. Unpublished Ph.D. thesis, Department of Statistics, North Carolina State University at Raleigh. University Microfilms, Ann Arbor, Michigan.

Bartle, R. G. 1964. The Elements of Real Analysis. John Wiley and Sons, Inc., New York, New York.

Goldberger, A. S. 1964. Econometric Theory. John Wiley and Sons, Inc., New York, New York.

Graybill, F. A. 1961. An Introduction to Linear Statistical Models. McGraw-Hill Book Company, New York, New York.

Gun, A. 1965. The use of a preliminary test for interactions in the estimation of factorial means. Unpublished Ph.D. thesis, Department of Statistics, North Carolina State University at Raleigh. University Microfilms, Ann Arbor, Michigan.

Johnson, N. L. and S. Kotz. 1970. Continuous Univariate Distributions. Houghton Mifflin, New York, New York.

Jordan, K. 1962. Calculus of Finite Differences. Chelsea Publishing Company, New York, New York.

Kempthorne, O. 1967. The Design and Analysis of Experiments. John Wiley and Sons, Inc., New York, New York.

Lancaster, P. 1969. Theory of Matrices. Academic Press, New York, New York.

Larson, H. J. and T. A. Bancroft. 1963. Biases in prediction by regression for certain incompletely specified models. Biometrika 50:391-402.

Rao, C. R. 1965. Linear Statistical Inference and Its Applications. John Wiley and Sons, Inc., New York, New York.

Sawa, T. and T. Hiromatsu. 1971. Minimax regret significance points for a preliminary test in regression analysis. Technical Report No. 39, The Economic Series, Stanford University, Stanford, California.

Theil, H. 1971. Principles of Econometrics. John Wiley and Sons, Inc., New York, New York.

Theil, H. and A. S. Goldberger. 1961. On pure and statistical estimation in economics. International Economic Review 2(1):102-126.

LIST OF REFERENCES (continued)

Toro-Vizcarrondo, C. E. and T. D. Wallace. 1968. A test of the mean square error criterion for restrictions in linear regression. Journal of the American Statistical Association 63:558-572.

Wilks, S. S. 1943. Mathematical Statistics. Princeton University Press, Princeton, New Jersey.